HEURISTI
SOLVIN

The abil

clusions while t

goal of medical

assumed that be

the Baconian id

facts which are

Systema

physicians, how

process of diag

tinually advanc

firmed. Physic

exclusively fo

and aggregatin

There

and conjecture

pretation incl

mation gatheri

ABSTRACT

HEURISTIC TRAINING FOR DIAGNOSTIC PROBLEM SOLVING AMONG ADVANCED MEDICAL STUDENTS

By

Michael Joseph Gordon

The ability to reach accurate diagnostic conclusions while treating patients humanely is a major goal of medical training. Medical schools have typically assumed that better diagnosis was to be achieved through the Baconian ideal of thorough and impartial gathering of facts which are later objectively interpreted and evaluated.

Systematic observation of competent practicing physicians, however, has led to the conclusion that the process of diagnosis is one in which hypotheses are continually advanced, tested, modified, ruled out, or confirmed. Physicians collect medical case data almost exclusively for the purposes of generating hypotheses and aggregating evidence in their favor.

There are obvious dangers in allowing hypotheses and conjectures to influence data collection and interpretation including premature closure, selective information gathering, and biased interpretation. Conversely,

there is reason
an indispensabl
the work-up. Th
direct the searc
greater economy
lection, hypoth
principles for
memory.

This st
of hypothesis-g
countered by s
instead by tra
help diagnosti
and to test th

A set
from analysis
diagnostic re
Thirty-two ad
Michigan medi
subjects.

In o
obtain evide
tic content
a series of
Half of the
mental heuri

there is reason to believe that these hypotheses may serve an indispensable function even in the earliest stages of the work-up. The formation of hypotheses appears to direct the search for information. In addition to the greater economy of focused rather than thorough data collection, hypotheses appear to function as the organizing principles for the storage and recall of information in memory.

This study has taken the position that the dangers of hypothesis-guided diagnostic inquiry should not be countered by struggles to eliminate early hypotheses, but instead by training in diagnostic heuristics which might help diagnosticians to generate more adequate hypotheses and to test their hypotheses more effectively.

A set of five experimental heuristics was derived from analysis of the reported and observed errors of diagnostic reasoning committed by medical students. Thirty-two advanced medical students attending two Michigan medical schools were selected as experimental subjects.

In order to test the thesis of this study and to obtain evidence of the effects of various kinds of heuristic content and usage, the students were presented with a series of medical cases which they were to diagnose. Half of the subjects were trained to employ the experimental heuristics and half were asked to generate and

employ a set of
they had found
solving. Withi
systematically
were invited t
All subjects i
solve the diag
accurately as
performance w
tic case. Th
(1) Scope of
the degree of
ses; (2) Num
the diagnost
financial ex
health inhe
(4) Accurac

The
sidered to
diagnostic
were consi
importance

Th
a set of d
important
tigated.

employ a set of personal or idiosyncratic heuristics which they had found to be helpful in past diagnostic problem solving. Within this division, half of the subjects were systematically prompted to use the heuristics and half were invited to use the heuristics at their own discretion. All subjects in the resulting four groups were asked to solve the diagnostic problems as efficiently and as accurately as possible. Four measures of problem-solving performance were taken for each subject on each diagnostic case. The dependent measures were defined as follows: (1) Scope of the early diagnostic formulations, reflecting the degree of generality or specificity of early hypotheses; (2) Number of critical findings elicited; (3) Cost of the diagnostic work-up, defined as an additive function of financial expense, patient discomfort, and risk to patient health inherent in the diagnostic procedures ordered; and (4) Accuracy of the diagnosis.

The Scope and Critical Findings measures were considered to be process measures which might be related to diagnostic outcomes. The measures of Cost and Accuracy were considered to be diagnostic outcomes of paramount importance.

The contribution of this study is twofold. First, a set of dependent measures for the quantification of important diagnostic outcomes has been defined and investigated. Those investigations demonstrated that the

evaluation of
be made objec
required to c
Second, the e
cess and outc
as relations
principal fi
were as foll
that the heu
performance
dependent m
the Accurac
nificance (
treatment e
be worthy
ment group
were more
(3) No sig
Scope of
Cost or t
Critical
both hig
relations
relation
Test Sco
school a

evaluation of diagnostic Cost and Accuracy performance can be made objectively but that several cases would be required to obtain acceptable coefficients of reliability. Second, the effects of problem-solving heuristics on process and outcome measures have been investigated as well as relationships among many performance variables. The principal findings resulting from these investigations were as follows: (1) There was no acceptable evidence that the heuristic training or prompting affected the performance of subjects on any of the four principal dependent measures; (2) Treatment group differences on the Accuracy measure approached acceptable levels of significance ($p < .07$). On this basis the hypothesis of treatment effects on the Accuracy measure is judged to be worthy of further pursuit. The trends between treatment groups suggested that the experimental heuristics were more beneficial than the idiosyncratic heuristics; (3) No significant relationship was found between the Scope of Early Diagnostic Formulations and either the Cost or the Accuracy of diagnosis; (4) The number of Critical Findings elicited was positively associated with both higher Cost and greater Accuracy, but no significant relationship was found between Cost and Accuracy; (5) No relationship was found between Medical College Admission Test Scores administered prior to entry into medical school and measures of diagnostic Cost or Accuracy.

The ge

training might

advanced medic

guided mode, h

Trends with re

Accuracy varia

favor of the

preted as fai

untested assu

medical educa

study indicat

history and

greater diag

greater diag

The general hypothesis of this study, that heuristic training might improve the problem-solving performance of advanced medical students functioning in a hypothesis-guided mode, has not been supported by the findings. Trends with respect to the group means on the diagnostic Accuracy variable, however, are encouraging evidence in favor of the hypothesis. The findings may also be interpreted as failing to support some of the previously untested assumptions of current pedigogical practice in medical education. Specifically, the results of this study indicate that greater thoroughness of the medical history and physical examination is associated with greater diagnostic Cost but is not associated with greater diagnostic Accuracy.

HEURIS

SOLVI

in

Depa

HEURISTIC TRAINING FOR DIAGNOSTIC PROBLEM SOLVING AMONG ADVANCED MEDICAL STUDENTS

By

Michael Joseph Gordon

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

1973

The co

me a very pers

proud to share

friends. I co

have had the

Craig, my aca

graduate work

Dr. Lee Shul

provided the

dissertation

role in the

Arthur Elste

enthusiasm

disappointm

from whom I

of good res

insights an

professed a

The

medical ca

ACKNOWLEDGMENTS

The completion of the doctoral dissertation is for me a very personal accomplishment, but one which I am proud to share with an exceptional group of teachers and friends. I consider myself to be extremely fortunate to have had the assistance of such people as Dr. Robert Craig, my academic advisor, who guided me through my graduate work in preparation for the dissertation; Dr. Lee Shulman, my dissertation director, who not only provided the guiding hand in the development of the dissertation but who also played a most significant role in the formation of my professional career; Dr. Arthur Elstein, whose intense curiosity and infectious enthusiasm often turned frustrations into challenges and disappointments into insights; and Dr. Andrew Porter, from whom I learned that good design is the _sine qua non_ of good research and who continually surprised me with insights and clarification in areas far removed from his professed area of expertise.

The construction and analysis of the simulated medical cases used in this study required the generous

assistance of
to Drs. Marvi
Gerald Holzma
Dr. Michael D

Of a
this accompl
than my wife
great sacrif
insistence t
balance wit

assistance of many physicians, but special thanks are due to Drs. Marvin Clark, Gilles Cormier, Donald Gragg, Gerald Holzman, Michael Spooner, and especially to Dr. Michael Doyle.

Of all of those with whom I would like to share this accomplishment, none has been more important to me than my wife, Katherine, whose faith in me and whose great sacrifices sustained my efforts; and at whose insistence the pursuit of academic goals was kept in balance with the larger goals of our life.

Chapter

I. BACKGR

The
The

The

Sta
Des

Id

E

II. REVI

D
T

TABLE OF CONTENTS

Chapter Page

I. BACKGROUND OF THE PROBLEM 1

The Stepwise Approach to Diagnosis . . . 1
The Hypothesis-Guided Approach to Diagnosis 5
The Disparity Between Training and Practice in Diagnostic Approach . . . 7
Statement of the Problem 10
Description and Rationale of Experimental Heuristics 11

1. Planning Heuristic 11
2. Hypothesis Specificity Heuristic . 12
3. Competing Hypothesis Heuristic . . 13
4. Re-interpretation Heuristic . . . 15
5. Negative Inference Heuristic. . . 15

Idiosyncratic Heuristics Versus the Experimentally Prescribed Heuristics . 17
Experimental Questions 20

II. REVIEW OF THE LITERATURE 24

Description of the Diagnostic Process . . 24
Theoretical Constructs Underlying Diagnostic Problem Solving. 36

Information Processing Theory. . . . 37
Decision Theory 43
Explanations of the Diagnostic Behavior of Physicians 46
Testing of Hypotheses 52

Errors in Diagnostic Reasoning 54

A. Failure of Planning. 56
B. Failures of Hypothesis Specificity 58

Chapter

The

III. PROCE

Su
De
Fo

I
C

IV. RE

Chapter Page

C. Failures to Devise Competing Hypotheses 59
D. Failures to Re-interpret Data . . 60
E. Failures of Negative Inference. . 61

The Effects and Effectiveness of Problem-Solving Heuristics 63

III. PROCEDURES 73

Subjects 73
Development of Diagnostic Cases. . . . 73
Format and Presentation of the Diagnostic Cases 77
Independent Variables 78
Construction and Analysis of Dependent Measures 82

Scope of Early Diagnostic Formulations. 85
Number of Critically Important Case Findings Elicited 89
Cost of Information Elicited in the Diagnostic Work-up 91
Accuracy of the Definitive Diagnosis . 96

Summary of Reliability Studies on Dependent Variables. 101
Statement of the Hypotheses 103
Experimental Design 104

IV. RESULTS. 108

Factorial Analysis of Dependent Measures 108

Scope of the Early Diagnostic Formulations (Scope) 108
Number of Critical Findings Elicited (Critical Findings). 109
Cost of Diagnostic Work-up (Cost) . . 111
Accuracy of the Definitive Diagnosis (Accuracy). 113

Correlational Analyses Among Dependent Variables 116

Chapter

I

V. DIS

VI. SU

SELECTED

Chapter Page

Consistency of Performance Across Problems 116
Relationship Between Diagnostic Cost and Accuracy. 119
Relationship Between Diagnostic Cost and Selected Process Measures of Problem Solving. 122
Relationship Between Diagnostic Accuracy and Selected Process Measures of Problem Solving. . . . 124
Relationships Between Process Measures 125

Incidental Analyses 127

Relationships Between Medical College Admissions Test Scores and Scores on Cost of the Diagnostic Work-up and Accuracy of the Definitive Diagnosis. 127
Evaluation of the Experimental Heuristics by Subjects 129
Content Analysis of Idiosyncratic Heuristics 130

Summary of Results. 132

V. DISCUSSION 137

Discussion of Schools Variable. 138
Discussion of Scope Results. 139
Discussion of Critical Findings Results. . 142
Discussion of Cost Results 145
Discussion of Accuracy Results. 149
Conjectures on Heuristic Processes and the Content of Medicine 153
The Relative Importance of Knowledge and Strategy 157

VI. SUMMARY AND CONCLUSIONS 159

Critical Review of the Procedures. . . . 165
Implications for Future Research and Development 169

SELECTED BIBLIOGRAPHY 177

APPENDICES

Appendix

A. Instruc
tra

B. Proble

C. Subjec

D. Supple

Page

APPENDICES

Appendix

A. Instructions, Training Notes, and Administrative Forms 183

B. Problem-Solving Case Data 190

C. Subject Scoring Forms 220

D. Supplemental Analyses 236

Table

1. Inter-
   Rat
   Dia
   Pro

2. Rounde
   ler
   Rat

3. Pears
   Or
   Co
   Pr

4. Corre
   Ac
   We

5. Trea
   o

6. Two-
   E

7. Trea
   c

8. Two-

9. Tre

10. Two

11. Tr

LIST OF TABLES

Table   Page

1. Inter-Rater Reliability of Two Independent Raters on the Measure, Scope of Early Diagnostic Formulations for Three Test Problems 89

2. Rounded Subjective Estimates of Cost-Equivalents for Selected Diagnostic Procedures Rated on Discomfort and Risk 94

3. Pearson Product Moment Correlations Between Original and Alternate Estimates of Total Cost Scores for Pretest and Posttest Problems 96

4. Correlations Among Scores of Diagnostic Accuracy Obtained by Four Systems of Weights 100

5. Treatment Group Means and Standard Deviations on Scope Measure 109

6. Two-Way Analysis of Covariance on Scope of Early Diagnostic Formulation Score 110

7. Treatment Group Means and Standard Deviations on Critical Findings Measure 110

8. Two-Way Analysis of Covariance on Number of Critical Findings Elicited 111

9. Treatment Group Means and Standard Deviations on Cost Measure. 112

10. Two-Way Analysis of Covariance on Cost of Diagnostic Work-up. 113

11. Treatment Group Means and Standard Deviations on Accuracy Measure 114

Table

12. Two-Way
the D

13. Pearson
form
Subje

14. Pearson
Cost
of D
All

15. Pearson
Cost
Proc
for

16. Pearson
Accu
Sele
Prob

17. Mean Pe
(Aft
Sele
for

18. Pearson
Scor
Test
on T
(N =

19. Pearso
Sco
Tes
Dia
Sub

20. Freque
Far
Tr
(N

21. Conte
by

Table Page

12. Two-Way Analysis of Covariance on Accuracy of the Definitive Diagnosis 114

13. Pearson Product Moment Correlations on Performance Measures Across Problems for All Subjects (N = 32) 118

14. Pearson Product Moment Correlations Between Cost of Diagnostic Work-up and Accuracy of Diagnosis on Individual Problems for All Subjects (N = 32) 121

15. Pearson Product Moment Correlations Between Cost of Diagnostic Work-up and Selected Process Measures on Individual Problems for All Subjects (N = 32). 123

16. Pearson Product Moment Correlations Between Accuracy of the Definitive Diagnosis and Selected Process Measures on Individual Problems for All Subjects (N = 32). 125

17. Mean Pearson Product Moment Correlations (After r to Z Transformation) Between Selected Process Measures Across Problems for All Subjects (N = 32) 126

18. Pearson Product Moment Correlations Between Scores on the Medical College Admissions Test and Cost of the Diagnostic Work-up on Three Problems for All Subjects (N = 32) 128

19. Pearson Product Moment Correlations Between Scores on the Medical College Admissions Test and Accuracy of the Definitive Diagnosis on Three Problems for All Subjects (N = 32) 129

20. Frequency of Subject Responses on Perceived Familiarity and Consistency with Clinical Training of Experimental Heuristics (N = 16) 131

21. Content of Idiosyncratic Heuristics Generated by Eight Subjects 133

Table

22. Proporti
Plus
ment

23. Verbatim
Subje

Table Page

22. Proportion of Critical Findings to Critical Plus Noncontributory Findings by Treatment Groups 145

23. Verbatim Idiosyncratic Heuristics of Group 3 Subjects 239

Figure

1. Plan of

2. Disease

3. Diagnos

4. Diagnos

5. Factori
   Gro
   of

6. Adjust
   Cos

7. Scatte
   Co

8. Scatt
   Co

9. Scatt
   C

LIST OF FIGURES

Figure Page

1. Plan of Experimental Procedures 83

2. Disease Process by Organ System Grid . . . 88

3. Diagnostic Summary Form 97

4. Diagnostic Accuracy Scoring Form. 98

5. Factorial Design for Analysis of Treatment Group and Medical School Effects on Each of Four Dependent Variables 105

6. Adjusted Mean Scores of Treatment Groups on Cost and Accuracy 120

7. Scatter Plot of Case 1 Correlation Between Cost and Accuracy 237

8. Scatter Plot of Case 3 Correlations Between Cost and Accuracy. 238

9. Scatter Plot of Case 4 Correlation Between Cost and Accuracy. 239

A

ing clin
pertinen
intellig
diagnos
virtual
procedu
(Harris
there
in whi
widely
is vi
first
histo
labo
in c
pos
of

CHAPTER I

BACKGROUND OF THE PROBLEM

The Stepwise Approach to Diagnosis

Among the necessary skills of a physician practicing clinical medicine are the abilities to collect the pertinent facts about a case and to use these facts intelligently in order to arrive at an appropriate diagnosis. Medical students develop these skills in virtually all U.S. medical schools through a set of procedures generally referred to as "the clinical method" (Harrison, 1970; Harvey & Bordley, 1966). Although there is some variation from school to school in the way in which the clinical method is taught, its outlines are widely agreed upon. In essence, the process of diagnosis is viewed as a sequential activity in which the clinician first collects data in the form of a thorough medical history and physical examination, and sometimes routine laboratory tests; then analyzes and synthesizes the data in order to reach a single diagnosis or a few diagnostic possibilities that can best account for the collection of data on hand. The physician may then seek further

data to suppor
between his re
activities imp
series of disc
diagnostic met
reference to

In th
diagnosis, me
great deal of
search for in
proper metho
largely unsp
diagnosis an
which will e
medical jud

Tex
systems or
the signs,
typical of
presented
brief ment
resemble
is expect
and in sc
in a clin
all other

data to support his hypothesis or to differentiate between his remaining hypotheses. This sequence of activities implies that diagnosis may proceed in a series of discrete steps. In subsequent discussions of diagnostic method the term "stepwise" will be used with reference to this approach.

In the teaching of the stepwise approach to diagnosis, medical schools have traditionally placed a great deal of emphasis on the thorough and systematic search for information but have left the details of proper methods of analysis and synthesis of the data largely unspecified. It is assumed that experience in diagnosis and native intelligence are the ingredients which will eventually produce the ability to make sound medical judgments.

Textbooks of diagnosis are arranged by organ systems or disease processes and provide discussions of the signs, symptoms, and abnormal laboratory findings typical of hundreds of different diseases; each disease presented in virtual isolation from all others with only brief mention of other disorders which are likely to resemble the disease in question. The medical student is expected to learn the manifestations of each disease and in some unspecified way, learn to apply his knowledge in a clinical setting--differentiating one disease from all others after he has collected a thorough list of

clinical fi
diagnosis,
from anothe
of diagnosi
eloquently
1966, p. 2)

> [Differ
> dangere
> It is,
> has be
> much o
> sympto
> To giv
> simula
> exceed
> retice
> thoug
> to ta
> diagn
> be of

A
applied
case fin
school c
of the r
among s
problem
availab
master-
collec
demon-
data-

clinical findings from the patient. Differential diagnosis, the process of distinguishing one disease from another, usually occupies one chapter in textbooks of diagnosis and is of questionable value for reasons eloquently stated by Richard Cabot (Harvey & Bordley, 1966, p. 2).

> [Differential diagnosis is] a very dangerous topic--dangerous to the reputation of physicians for wisdom. It is, I suppose, owing to this danger that so little has been written on differential diagnosis and so much on diagnosis (nondifferential). To state the symptoms of typhoid perforation is not difficult. To give a set of rules whereby the conditions which simulate typhoid perforation may be excluded is exceedingly difficult. Physicians are very naturally reticent on such matters, slow to commit their thoughts to paper, and very suspicious of any attempt to tabulate their methods of reasoning. Yet all diagnosis must become differential before it can be of any use.

Although the methods of reasoning that should be applied in the analysis and synthesis of diagnostic case findings have not been a welcome subject in medical school curricula, medical educators have been well aware of the manifestations of faulty diagnostic problem solving among students. They have sought the cure for inadequate problem solving, however, in the two areas that are most available to observation and remediation; greater mastery of medical knowledge and more thorough data collection in the case at hand. The student who has demonstrated mastery of medical content and compulsive data-gathering habits, but who still makes errors of

diagnostic j
members, wh
with experi

The
of thorough
analysis ar
insisted up
the most s
These erro
of insuffi
informati
accompani
that may
the physi
of findi

of the e
separati
medical
is comp
ment of
clinica
(Elstei
& Simon
of prac
and doc

diagnostic judgment is an enigma to medical faculty members, who can only hope that the student will improve with experience.

The almost universally endorsed stepwise method of thorough, systematic data collection followed by analysis and synthesis of the findings is firmly insisted upon in order to minimize the likelihood of the most salient errors of diagnostic problem solving. These errors are (a) leaping to conclusions on the basis of insufficient evidence, (b) selective elicitation of information to confirm a favored hypothesis (often accompanied by verbal inflections and nonverbal cues that may dispose a patient to provide the answer that the physician expects), and (c) biased interpretations of findings toward the physician's favored hypothesis.

Considering the frequency and potential severity of the errors of judgment mentioned, the rationale for separating the systematic and objective elicitation of medical data from the analysis and synthesis of the data is compelling. But despite the near unanimous endorsement of the stepwise procedures and rationale of the clinical method, recent studies of practicing physicians (Elstein, Kagan, Shulman, Jason, & Loupe, 1972; Schwartz & Simon, 1970) demonstrate that the overwhelming majority of practicing physicians, including academic physicians and doctors considered to be excellent diagnosticians by

their colleag
when confront
tic behavior
be labeled th

In t
from observa
sicians, a f
generated a
of the pati
or specific
of the diag
hypotheses
further sy
sician is
might tend
entertain
is freque
the desir
risking t

I
collectio
condition

1.

their colleagues, actually perform quite differently when confronted with a diagnostic problem. The diagnostic behavior of competent, experienced physicians might be labeled the "hypothesis-guided" approach to diagnosis.

The Hypothesis-Guided Approach to Diagnosis

In the hypothesis-guided method, as inferred from observations and interviews with experienced physicians, a few tentative diagnostic possibilities are generated after only brief observation and questioning of the patient. These hypotheses may be either general or specific and usually emerge in the first few minutes of the diagnostic work-up. Instead of keeping these hypotheses "in the back of his head" and proceeding with further systematic (routine) data collection, the physician is guided by his hypotheses to elicit data which might tend to confirm or disconfirm the diagnoses he is entertaining. Routine (nonhypothesis-guided) information is frequently elicited, but serves functions other than the desire for thoroughness of the data base prior to risking the formulation of hypotheses.

In the hypothesis-guided method, routine data collection is pursued under any of the following three conditions:

1. When the physician has exhausted his current diagnostic hypotheses through testing, he will

reso

enou

new

2. When

he n

hypo

hypo

to h

dia

3. Aft

the

sys

hi

as

f

n

m

r

to as e

study t

"stepw

ventio

resort to routine data collection until he has enough new data to generate and test one or more new hypotheses.

2. When the physician has collected information that he needs to sort out with respect to his current hypotheses or the possible incorporation of new hypotheses, he may ask for routine data in order to have time to think about and organize his diagnostic problem.

3. After a satisfactory diagnosis has been reached, the physician may conclude the work-up with a systematic functional inquiry in order to assure himself that he has not overlooked any important aspect of the case. This inquiry is viewed as a fail-safe procedure in which the physician expects no findings inconsistent with his diagnosis but may occasionally turn up evidence of an unrelated medical problem, an unsuspected complication or in rare instances some information which requires a reconsideration of the diagnosis.

Reasoning processes are more typically referred to as either inductive or deductive. In the present study the author has chosen to use the descriptive terms "stepwise" and "hypothesis-guided" rather than the conventional terms because in each of these diagnostic

approaches, b
though with d
diagnostic pr
deductive de
to be genera
and the hypo
consistent w

The
students ar
practicing
One questi
is how out
the stepw
This ques
Sprafka (
permitted
often mak
premature
and bias

presentl
data pri
into pra

approaches, both inductive and deductive reasoning occurs, though with different emphasis at various stages of the diagnostic process. Readers who prefer the inductive-deductive designation may consider the stepwise approach to be generally more consistent with inductive reasoning and the hypothesis-guided approach to be generally more consistent with deductive reasoning.

The Disparity Between Training and Practice in Diagnostic Approach

The disparity between the way in which medical students are taught to perform a diagnosis and the way practicing physicians perform raises a number of questions. One question of central importance for medical education is how outcomes of diagnostic problem solving differ under the stepwise method and the hypothesis-guided method. This question has been under investigation in part by Sprafka (1973). It may be that medical students, if permitted to use the hypothesis-guided method, will more often make diagnostic errors due to tendencies toward premature conclusions, selective collection of information, and biased interpretations of findings.

On the other hand, it is realized that students presently trained in methods of thorough collection of data prior to analysis will, a short time after entry into practice, begin operating in the hypothesis-guided mode. Among

shift in beh

1. The
   prac
   coll
   exam
   univ
   may
2. The
   cost
   mone
   of
   to
   hos
3. The
   of
   com
   Dri
   muc
   tiv
   Th
   in
   in
   pr

mode. Among the probable reasons for the predictable shift in behavior are the following:

1. The time constraints of the typical office practice usually do not permit the extensive collection of medical history and physical examination data that are collected in a university hospital; so the busy physician may begin to take short-cuts.
2. The emphasis on thoroughness may increase the cost of medical care to patients in terms of money, discomfort, and risk to health. Each of these concerns becomes far more important to a physician when he leaves the university hospital and enters private practice.
3. The emphasis on thoroughness ignores the problem of information overload. Studies of cognitive complexity and information processing (Schroder, Driver, & Streufert, 1967) indicate that too much information may be as detrimental to effective problem solving as too little information. The hypothesis-guided method may reduce the information overload in two ways: by eliminating information not considered to have a high probability of contributing to the solution of

the

"chu

unde

The

ficient rea

stepwise me

often insu

usually en

physicians

even if th

method to

T

sicians a

approach

taneousl

approach

ciency a

to orga

informa

tic er

interp

diagn

regar

cessi

hypo

the diagnostic problem and by permitting the "chunking" and the organization of information under tentative diagnostic rubrics.

The problem of information overload may be sufficient reason in itself to force the abandonment of the stepwise method. Physicians in university hospitals are often insulated from many of the time and expense concerns usually encountered in private practice. Yet academic physicians typically employ the hypothesis-guided method, even if they also happen to be teaching the stepwise method to medical students!

Thus it appears that although virtually all physicians are thoroughly trained to use one kind of approach to diagnostic problem solving, they spontaneously combine it with a second approach. The latter approach offers the assumed advantages of greater efficiency and greater compatibility with the human capacity to organize, store, retrieve, and otherwise manipulate information, but introduces increased dangers of diagnostic errors through selective search procedures, biased interpretations, and premature conclusions.

It is the thesis of this study that training in diagnostic problem solving should proceed with due regard to the limitations on human information processing. Such training would follow the lines of a hypothesis-guided approach. It is further proposed that

methods o
diagnosti
improve t
efficient
effective
tic judg

of diag
mastery
of case
Methods
systema
expecte
activi
clinic
ings a
autops
are d

ing i
diagn
using
trai
the

methods of training in general heuristic rules of diagnostic problem solving can be devised which will improve the skills of medical students in conducting efficient information search and in using information effectively and prudently to arrive at accurate diagnostic judgments.

Statement of the Problem

As previously mentioned, the emphasis in training of diagnostic skills has traditionally been placed on the mastery of medical knowledge and the thorough collection of case data preceding attempts at diagnostic reasoning. Methods of diagnostic reasoning per se have not been systematically trained. Instead, medical students are expected to acquire reasoning skills through such activities as supervised clinical experience and clinical-pathological conferences, in which case findings are presented in detail--usually from birth to autopsy--and the diagnostic and treatment implications are discussed.

This study has attempted to determine how training in rules of diagnostic reasoning might affect the diagnostic performance of advanced medical students using the hypothesis-guided diagnostic approach. The training employed was meant to provide the student with the ability to use a set of five heuristics, or rules of

thumb, whic
in virtuall
ristics were
aid the stu
gathered an
the differe

The
problem-sol
from previo
medical and
studies of
analysis o
physicians
five heuri
of this st
of each he
cussion o

1. Plann

solver s
the prob
purpose

divided

thumb, which were sufficiently general to be applicable in virtually all areas of clinical diagnosis. The heuristics were intended as self-monitoring checks which aid the student in the efficient selection of data to be gathered and in the effective use of his data base in the differential diagnosis.

Description and Rationale of Experimental Heuristics

The set of experimental heuristics used as problem-solving aids for this study have been derived from previous studies of the problem-solving process in medical and nonmedical domains, empirical and theoretical studies of heuristic methods, and from observation and analysis of diagnostic performance among experienced physicians. The literature supporting the choice of the five heuristics selected will be reviewed in Chapter II of this study. In the following paragraphs a description of each heuristic rule is presented with a brief discussion of the rationale for its inclusion.

1. Planning Heuristic

Each piece of information requested by the problem solver should be related to a plan of attack for solving the problem. There should be a plan and a well-defined purpose behind every question asked.

Every nontrivial diagnostic problem can be divided into an organized set of subproblems. In most cases the se

has a direct

of the diag

to plan are

First, in a

information

students i

of data, t

data leadi

diagnostic

focus pre

gated mea

been reso

the commo

tory test

establis

question

2. Hypo
Heuristi

or more

a parti

"proble

find a

the di

cases the sequence in which the subproblems are addressed has a direct bearing on the efficiency and effectiveness of the diagnostic inquiry. Two manifestations of failure to plan are commonly observed among medical students. First, in a "dragnet" approach, some students will elicit information exhaustively and indiscriminately. Such students implicitly assume that through sheer quantity of data, they will eventually uncover the appropriate data leading to the diagnosis. Second, an unplanned diagnostic inquiry frequently leads medical students to focus prematurely on a subproblem that cannot be investigated meaningfully until prior medical questions have been resolved. For example, medical educators refer to the common student practice of ordering expensive laboratory tests to confirm a diagnosis that could have been established or ruled out with a few medical history questions or a brief physical examination.

2. Hypothesis Specificity Heuristic

No diagnostic hypothesis should be more specific or more general than the evidence on hand justifies.

The diagnostic hypotheses under consideration at a particular time help determine the psychological "problem space" in which the diagnostician expects to find a solution. The data elicited become evidence for the diagnoses being considered. Therefore, when

hypotheses
problem spa
reduced, an
the space i
than the ev
becomes lar
required to
(1972) repo
is to use a
pain and su
such as ang
rubric (e.g
angina pect
possible sc
an unwillin
justifiabl
hypotheses
size and c
tually eli
diac abnor
limit the
congenital
other chro

3. Compet
Heuristic

Th
competing

hypotheses are unjustifiably specific the size of the problem space in which the problem solver operates is reduced, and the likelihood of finding the solution within the space is less. When the hypotheses are more general than the evidence on hand justifies, the problem space becomes larger than necessary, and the time and effort required to search it is increased. Barrows and Bennett (1972) report that a common tendency of medical students is to use a general patient's complaint such as chest pain and suggest a very specific diagnostic hypothesis such as angina pectoris, rather than a more general rubric (e.g., cardiovascular disease) that would include angina pectoris as well as a large number of other possible solutions. On the other hand some students show an unwillingness to use the data on hand which might justifiably allow them to limit the scope of their hypotheses. For example, a finding of normal heart size and contour should permit the clinician to virtually eliminate the possibility of long-standing cardiac abnormality. When such findings are not used to limit the scope of hypotheses an unproductive search for congenital abnormalities, rheumatic heart disease, and other chronic conditions may needlessly follow.

3. Competing Hypothesis Heuristic

There should always be at least two or three competing hypotheses under consideration at a particular

time. Each
with respect
sideration.

Duri
several stuc
early in the
confirm the
faculty mem
toward prem
common. Ma
their level
reduce the
situation b
ically the
the importa
ambiguous f
direction.
of hypothe
part by th
ing hypoth
dence can
make it ea
hypothesis
breadth of
solver to

time. Each piece of information should be evaluated with respect to all hypotheses presently under consideration.

During the pilot testing phase of this study, several students appeared to leap to a single hypothesis early in the work-up and to seek information intended to confirm the favored hypothesis exclusively. Clinical faculty members interviewed about student tendencies toward premature closure agreed that the error was common. Many medical students who feel insecure about their level of diagnostic skill apparently attempt to reduce the anxiety accompanying an ambiguous clinical situation by "forcing a diagnosis" on the patient. Typically the importance of negative evidence is minimized, the importance of positive evidence is magnified, and ambiguous findings are interpreted in a positive direction. Such "case building" is an obvious hazard of hypothesis-guided inquiry, but can be controlled in part by the insistence on at least two or three competing hypotheses (Chamberlin, 1965) against which the evidence can be evaluated. Competing hypotheses may also make it easier for the problem solver to discard a hypothesis found to be unfruitful or to increase the breadth of the problem space available for the problem solver to work in.

4. Re-inter

Wher
information
problem sol
elicited fi
to disconfi

New
dent with a
study it wa
students to
the new hyp
on hand whi
Kleinmuntz
reasoning,
the time t
the soluti
reconsider
to bring a
sis and a

5. Negati

Wh
risky) pro
favored hy
the possib
instead ru
in order t

4. Re-interpretation Heuristic

Whenever a new or revised hypothesis emerges, the information previously collected should be reviewed. The problem solver should attempt to categorize the previously elicited findings as either tending to confirm or tending to disconfirm his new hypothesis.

New or revised hypotheses usually emerge coincident with a new finding. In the pilot testing of this study it was noted that there was a tendency among medical students to proceed to collect more evidence related to the new hypothesis and to disregard the findings already . on hand which might help to confirm or disconfirm it. Kleinmuntz (1968), on the basis of his studies of medical reasoning, made the claim that evidence collected prior to the time that a hypothesis was formulated was not used in the solution of a diagnostic problem. A brief effort at reconsideration of data previously collected is expected to bring about a more accurate assessment of the hypothesis and a more efficient search for further data.

5. Negative Inference Heuristic

When high cost (expensive, uncomfortable, or risky) procedures are being considered to confirm a favored hypothesis, the problem solver should consider the possibility of lower cost procedures which might instead rule out one or more diagnostic possibilities in order to make the high cost procedure unnecessary or

to increase
will yield t

A ps
solving stud
might confir
would be mor
favored hyp
ses. An ex
showed up i
ject had na
alternative
surgical bi
confirmatio
a simple ur
aware of th
he chose to
firmatory s
less costly

In
their use
presented
by the sub
responding
given as f

1. Wh
es

to increase the probability that the high cost procedure will yield the definitive diagnosis.

A psychological set uncovered in many problem-solving studies is the tendency to look for evidence which might confirm the most favored hypothesis, even when it would be more efficient to try either to disconfirm the favored hypothesis or to confirm the less likely hypotheses. An excellent example of this kind of behavior showed up in the pilot testing of this study. The subject had narrowed his diagnostic hypotheses to three alternatives. His most favored hypothesis required a surgical biopsy involving some risk to the patient for confirmation. His second-ranked hypothesis required only a simple urine test for confirmation. Although he was aware of the urine test and had specifically mentioned it, he chose to perform the biopsy first. The subject's confirmatory set apparently precluded his selection of the less costly but equally diagnostic procedure.

In order to aid the subjects of the study in their use of the heuristic rules, each rule was also presented in the form of a question to be considered by the subject. The five heuristic questions corresponding to each of the five heuristic rules were given as follows:

1. What is it that you would like to accomplish or establish with the next set of questions you ask?

2. Is your hypothesis at an appropriate level of specificity or generality given what you know about the case so far?
3. Given the information you have so far, are there any other hypotheses that you should be considering? Have you tested the last group of findings against all of your hypotheses?
4. Can you think back over your findings and try to find the pieces of information which tend to confirm or tend to disconfirm your new hypothesis?
5. Before you perform a high-cost procedure to confirm a hypothesis, can you think of a low-cost procedure which might instead rule out one or more of your hypotheses--or at least demonstrate more clearly the need to perform the high-cost procedure?

Idiosyncratic Heuristics Versus the Experimentally Prescribed Heuristics

As Polya (1957) has observed, all sane people use heuristics to solve problems. Medical students are expected to learn, through clinical experience, habits and methods of clinical inquiry which include the development of a set of natural heuristics for solving clinical problems. It may be that these natural or idiosyncratic heuristics are more beneficial to medical

students
they hav
experien
quently

to be e
methods
less th
golfer
grip an
gested
golfer
new for
and aut

(1965)
is more
mediat
to fav
Even i
ever,
of med
mnemon
is enh
Specif
are hi

students than the set of experimental heuristics because they have been developed from the student's own unique experience and from the analysis of his own most frequently made or most costly errors.

On the other hand not all heuristics are expected to be equally effective, and the naturally developed methods of clinical thought and procedure may be far less than optimal. To draw an analogy, the self-taught golfer may feel very comfortable with his idiosyncratic grip and swing and very awkward using a new form suggested by a professional. But in the long run the golfer will probably be better off accommodating to the new form and practicing it until it too becomes natural and automatic.

Research by Reese (1965) and Jensen and Rohwer (1965) has demonstrated that paired-associate learning is more effective using one's own verbal mediators than mediators provided by others. This finding would seem to favor, by analogy, the use of idiosyncratic heuristics. Even in the paradigm of paired-associate learning, however, it would seem that external rules for the choosing of mediators can be helpful. The experience of stage mnemonicists indicates that paired-associate learning is enhanced by the use of particular kinds of mediators. Specifically, mediators conveying visual images which are highly dynamic, concrete, over-sized, or ridiculous

are more effe
abstract, or
is reason to
associate le
be self-gene
nally provid

In c
heuristics
experimenta
Perhaps the
the specifi
instead li
to re-cons
to raise t
the proble
tics may
given any
thoroughl

along tw
Subjects
experime
subject
continu
to use
Compar

are more effective than mediators which are static, abstract, or reasonable (Lorayne, 1957). Thus there is reason to believe that although mediators in paired-associate learning and specific diagnostic inquiry should be self-generated, both processes may be aided by externally provided guides.

In considering the question of which kind of heuristics are more effective--the idiosyncratic or the experimentally prescribed--there is a third possibility. Perhaps the value of a heuristic does not derive from the specific mental operations that it calls for, but instead lies in the motivation to reorganize the problem, to re-consider the data, to persevere, or in any other way to raise the general level of mental activity regarding the problem. Thus the important aspect of using heuristics may not be the quality of the heuristics but instead--given any set of heuristics--how systematically and thoroughly they are applied.

In this study the use of heuristics was varied along two dimensions in order to address these questions. Subjects used either their idiosyncratic heuristics or the experimental heuristics. Within this division, half of the subjects were given prompts to aid their systematic and continuous use of the heuristics while half were free to use the heuristics at their disposal as they wished. Comparisons among the four resulting groups were made

to shed s
simply st
systemati
employmen

I

of heuris
to test
versus t
students
cases.
treatmen
and rece
training
tic prob

heurist
part on
tent an
for the
trainin
drawn i
Univers
The Uni
regarde

to shed some light on a question which might be most simply stated as follows: Which is more important, the systematic employment of any set of heuristics or the employment of heuristics of a particular kind?

Experimental Questions

In order to determine the possible effectiveness of heuristic training in diagnostic problem solving and to test the relative effectiveness of the experimental versus the idiosyncratic heuristics, advanced medical students were presented with a series of diagnostic cases. Each student was assigned to one of the four treatment groups described in the preceding section and received a prescribed combination of heuristic training and prompting in conjunction with the diagnostic problem-solving tasks.

The ability to profit from one's idiosyncratic heuristics or the experimental heuristics may depend in part on the subject's previous training in medical content and diagnostic problem solving. In order to account for the effects of possible differences in previous training, the subjects in each treatment group were drawn in equal numbers from the medical schools of the University of Michigan and of Michigan State University. The University of Michigan is a well-established, highly regarded institution in which clinical training is

accomplished p
two nearby hos
to the Univers
school is cons
sidered to be
taking place p
with greater a

All s
hypotheses an
efficiently as
solving perfor
each diagnosti
follows:

(1) Scope
(2) Numbe
(3) Cost
(4) Accur

Measu
by researcher
tic problem s
to be outcome
diagnostic pr
the dependent
Chapter III o

The m
follows:

accomplished primarily in a University hospital and in two nearby hospitals with long-standing academic ties to the University. The Michigan State University medical school is considerably smaller and newer, generally considered to be more innovative, with clinical training taking place primarily in outlying community hospitals with greater autonomy from the University.

All subjects were instructed to follow their hypotheses and to search for an accurate diagnosis as efficiently as possible. Four measures of problem-solving performance were taken from each subject on each diagnostic case. The dependent measures were as follows:

(1) Scope of the Early Diagnostic Formulations
(2) Number of Critical Findings Elicited
(3) Cost of the Diagnostic Work-up
(4) Accuracy of the Diagnosis

Measures 1 and 2 have been previously identified by researchers as important process variables in diagnostic problem solving. Measures 3 and 4 were considered to be outcome variables by which the effectiveness of diagnostic problem solving could be judged. Each of the dependent measures is extensively described in Chapter III of this study.

The main research questions of the study are as follows:

1. Do adv
combin
perfor
measu

2. Do sc
at th
of ad
State
meas

3. Are
of h
medi
depe
adva

4. Are
obt
den

No
ects used
made that f
a rational
of implicit
the problem
extent of
involving

1. Do advanced medical students receiving different combinations of heuristic training and prompting perform differently on any of the four dependent measures defined in this study?
2. Do scores of advanced medical students enrolled at the University of Michigan differ from scores of advanced medical students enrolled at Michigan State University on any of the four dependent measures defined in this study?
3. Are there interactions between the combination of heuristic training and prompting and the medical school of enrollment on any of the four dependent variables defined in this study for advanced medical students?
4. Are there significant relationships among scores obtained by advanced students on the four dependent measures?

No specific measures of the extent to which subjects used heuristics were obtained. The assumption was made that for all subjects, diagnostic inquiry would be a rational rather than rote process in which some kind of implicit or explicit rules of thumb would be guiding the problem-solving effort. Therefore the question of extent of heuristic use was secondary to the questions involving the type of heuristic used (experimental or

idiosyncrati
particular
dimensions
manipulated
mental gro
requiring
select an
an answer

T
mally sta
of the nu
sidiary
related
also rep

idiosyncratic) and how systematically a small set of particular heuristics was employed. Both of these dimensions of heuristic use were experimentally manipulated (1) by exposing only particular experimental groups to the experimental heuristics and (2) by requiring particular experimental groups regularly to select an appropriate heuristic question and to verbalize an answer to it.

The foregoing research questions are more formally stated as null hypotheses in Chapter III. Tests of the null hypotheses are reported in Chapter IV. Subsidiary research questions and additional findings related to diagnostic problem-solving performance are also reported in Chapter IV.

T
each foc
study.
physicia
Second,
behavior
decision
literatu
suggesti
Fourth,
of probl

investig
avenue o
mathemat
Bayesian
attempti

CHAPTER II

REVIEW OF THE LITERATURE

This review will be divided into several sections, each focusing on a separate topic relevant to the present study. First, the literature describing the behavior of physicians in diagnostic settings will be surveyed. Second, explanations of diagnostic problem-solving behavior derived from information processing theory and decision theory will be examined. Third, the available literature on common diagnostic reasoning errors and suggestions for reducing these errors will be reviewed. Fourth, theoretical and empirical works in the domain of problem-solving heuristics will be discussed.

Description of the Diagnostic Process

There have been two main thrusts in the recent investigations of the diagnostic process. In the first avenue of research, investigators are working to develop mathematical models of decision making. Using primarily Bayesian and regression techniques these authors are attempting to devise algorithms for manipulation of clinical

There has

processes

data are

and the

function

as accura

have been

1960) of

attempts

judges.

I

are atte

reasoning

researche

jects in

relying

"thinkin

processe

these in

referrin

of human

Process

vations

vations

clinical findings which will yield optimal diagnoses. There has been no attempt to simulate human judgment processes per se in this research. Rather, the clinical data are seen as input; the diagnosis is seen as output; and the task is to define a parsimonious mathematical function which will map known input to criterial output as accurately as possible. These mathematical functions have been termed paramorphic representations (Hoffman, 1960) of human judgment and should be distinguished from attempts to model the actual reasoning processes of human judges.

In the second avenue of research, investigators are attempting to describe and account for the actual reasoning processes of competent physicians. These researchers are studying the behavior of physician subjects in various kinds of simulated clinical settings relying on a combination of observational techniques and "thinking aloud" reports of the physicians' reasoning processes as a data base. The present study follows from these investigations. Therefore, only the literature referring to these so-called isomorphic representations of human judgment will be reviewed.

The existing characterizations of the diagnostic process have been built on necessarily limited observations by relatively few investigators. Their observations have been made using different settings, different

problem co
Rather tha
researcher
of this re
possible,
descripti
ing the a
and varia
diagnosti
tive mode
ing this
turn to a
observati

T
patient
A number
moments
vations.
process
problema
the phys
that wa
of the
and Ben
medical
immedia

problem content, different formats and instructions. Rather than compare the theoretical models built by each researcher to explain his observations, the first part of this review will limit its discussion, as much as is possible, to the observations reported and to the summary descriptions inferred from observations. After clarifying the areas of general agreement, areas of controversy, and variations of emphasis with respect to aspects of diagnostic behavior, the outlines of a probable descriptive model of diagnostic behavior will be defined. Following this tentative descriptive model, the review will turn to a discussion of the theories underlying the observations.

The diagnostic encounter typically begins with a patient coming to a doctor with one or more complaints. A number of investigators have focused on the first few moments of this encounter and have made similar observations. Elstein et al. (1972) have noted an initial process of "cue attendance" by the physician from which problematic elements emerge. Wortman (1972) described the physician as extracting pertinent medical information that was easily detected upon initial cursory examination of the patient. Schwartz and Simon (1970) and Barrows and Bennett (1972) add that in addition to strictly medical signs such as pallor or fatigue, physicians immediately note several additional characteristics

including ap
patterns. F
physician be
about a cas
view.

Imr
with the n
to be made
long-term
as "associ
cues and
which res
Schwartz
that, "At
patient
edge, he
level of
cluded t
hypothe
clinici

eration
Shulman
Schwart
functio
is rem

including approximate age, sex, race, dress, and behavior patterns. From these observations it appears that the physician begins actively collecting and organizing data about a case from the moment that the patient comes into view.

Immediately following or perhaps simultaneously with the noting of these initial cues, connections appear to be made between the salient cues and the physician's long-term memory. Elstein et al. (1972) reported these as "associations between the problematic elements of the cues and the physician's store of medical knowledge," which results in the formulation of diagnostic hypotheses. Schwartz and Simon (1970) confirm this observation noting that, "At this point on the basis of initial complaint, patient contact, and the physician's structure of knowledge, he already begins to formulate hypotheses at some level of specificity." Barrows and Bennett (1972) concluded that among the neurologists they have studied hypotheses appear to literally pop into the head of the clinician "almost before the interview begins" [p. 275].

Whether the mechanism of early hypothesis generation is an associative one as implied by Elstein and Shulman, an information processing routine suggested by Schwartz and Simon or perhaps a pattern recognition function suggested by Lusted (1971) is not clear. What is remarkable and well established is the fact that

physicians o
hypotheses r

The
been disput
reported th
a series of
format beg
narrowed t
strategy i
questions"
studying
lation, a
clinician
overlapp
shape the
down" st

approach
tasks a
general
stage t
This c
decisi
These
approp
cues t

physicians operating on a very meager data base develop hypotheses rapidly, naturally, and virtually automatically.

The specificity of the earliest hypotheses has been disputed among researchers. Kleinmuntz (1968) reported that a well-known clinical neurologist solving a series of neurological problems in a "20 questions" format began with very general hypotheses and successively narrowed their scope. Of course, a general to specific strategy is the approach par excellence for solving "20 questions" types of problems. Barrows and Bennett (1972), studying clinical neurologists in a more realistic simulation, also noted, however, that the more experienced clinicians tended to begin with vague, general, somewhat overlapping hypotheses and to successively narrow and shape these hypotheses in what they called a "coning down" strategy.

Wortman (1972) analyzed the hypothesis-formulation approach of Kleinmuntz's subjects in simulated diagnostic tasks and concluded that hypotheses progressed from general to specific, but commenced at the most specific stage that was reasonable, given the data available. This conceptualization is in partial accord with a decision model put forth by Schwartz and Simon (1970). These authors contend that the physician decides upon an appropriate level of specificity by referring the initial cues to his hierarchically organized store of medical

knowledge.
which his da
data to diff
sician succe
lations wit
a match bet
general for
differenti
physicians
working fr
sicians w
zations,
located,
and narro

to speci
et al.
physici
hypothe
generat
may be
limited
simula
of com
to det
diagno

knowledge. He first looks for specific diseases between which his data might differentiate. Having insufficient data to differentiate between specific diseases, the physician successively looks for more and more general formulations within his organization of medical knowledge until a match between the available cues and a group of more general formulations of the problem provides a basis for differentiation among hypotheses. Thus for Wortman, physicians find the appropriate level of generality by working from the top down; for Schwartz and Simon, physicians work from the bottom up. In both conceptualizations, once the appropriate level of generality is located, the physician attempts to differentiate further and narrow the limits of his hypotheses.

In contrast to the many observations of a general to specific hypothesis formulation approach, Elstein et al. (1972) have noted that many of their experienced physician subjects often entertained quite specific hypotheses early on. The tendency of these subjects to generate specific rather than general early hypotheses may be due to a number of factors including but not limited to instructions, area of medicine, reality of the simulation, physician style, or, as Barrows implies, level of competence. Further work in this area seems indicated to determine both normative performance and import for diagnostic competence.

A
is the nu
Eistein e
that rega
quantity
the simul
physicia
simultan
than thr
that stu
crete ob
taneousl
It is re
narrow l
hold co
Related
physics
within
Miller'
chologi
the ph
will b
on inf

If he
these

Another intriguing aspect of hypothesis generation is the number of hypotheses entertained at any one time. Elstein et al (1972) and Barrows and Bennett (1972) concur that regardless of the specificity of the hypotheses, the quantity or quality of the data available, the format of the simulation task, the expertise or experience of the physician, or the content area of medicine, the number of simultaneously entertained hypotheses is seldom fewer than three or greater than five. Wortman (1972) reported that students solving diagnosis-type problems using concrete objects of varying size, shape, and color simultaneously entertained between three and seven hypotheses. It is remarkable that no marked deviations from these narrow limits have been reported and that these limits hold consistently across such a wide variety of situations. Related findings in fields of study as diverse as psychophysics and verbal memory have generated much discussion within the context of information processing theory. Miller's provocative 1956 address to the American Psychological Association stands as a classic description of the phenomenon and a plausible explanation. This point will be discussed in great detail in a subsequent section on information processing theory.

What does the physician do with his hypotheses? If he were to follow the advice of his medical texts these hypotheses would be at least temporarily suppressed.

The editor
(1970), fo
sicians ba
stress tha
avoid the
irritabil
thorough
he would
a differe
nostic c
hensive
a repres
writers

1.

2.

exis

The editors of Harrison's Principles of Internal Medicine (1970), for example, acknowledge the "prejudice" of physicians based upon previous medical experience. They stress that the physician "must struggle constantly to avoid the bias occasioned by his own attitude, mood, irritability and interest" until the completion of a thorough history and physical examination. At this point he would be free to synthesize his data base, construct a differential diagnosis and a plan for reaching a diagnostic conclusion. Harvey and Bordley, in their comprehensive volume on Differential Diagnosis (1966), provide a representative statement of what are seen by textbook writers to be the successive steps leading to a diagnosis:

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.
    - (f) Review all the evidence--both positive and negative--with the final diagnosis in mind.

A substantial and growing amount of research exists to challenge this pedagogical advice. Although

the advice a
not to foll
diagnostic
found that
tinuous hyp
arrive at

El
sician sub
workup not
to riskin
hypothese
patient"
". . .
time. T
mation,
[p. 8].
continu
is the
guides
Kleinmu
regard
series
90% of
hypoth
purpos
or si

the advice appears logical, experienced physicians tend not to follow it. Beginning with observations of the diagnostic process by Jacquez (1964) investigators have found that physicians employ a universal mode of continuous hypothesis formulation and testing in order to arrive at diagnoses.

Elstein (1972) reported that his typical physician subject utilized the familiar systematic diagnostic workup not to constrain a universe of possibilities prior to risking a formulation but instead "to test specific hypotheses he has formed early in his contact with the patient" [p. 11]. Schwartz and Simon (1970) reported " . . . judgments and decisions are occurring all the time. The physician does not passively gather information, stop, synthesize it all and reach a diagnosis" [p. 8]. Rather he is engaged in an iterative process of continuously reformulating and testing hypotheses. It is the current set of hypotheses, they concluded, which guides the physician's diagnostic behavior. Wortman and Kleinmuntz (1972) found that a well-known and highly regarded clinical neurologist who was asked to solve a series of neurological problems devoted between 78% and 90% of his verbal behavior to the testing of diagnostic hypotheses. The remaining verbalizations were for the purpose of generating hypotheses. (There was no actual or simulated patient present in this study. In an actual physician-p

purpose of

would be e

cluded tha

presented

only igno

hamper it

I

engaged

or less

patient.

illness

history

from he

tests.

sisten

ing.

the sy

the pu

enter

physi

routi

sici

at s

scre

or

physician-patient encounter some verbalizations for the purpose of establishing rapport or giving instructions would be expected.) Barrows and Bennett (1972) concluded that the standard steps of diagnostic medicine as presented in textbooks and emphasized in training not only ignore the discipline of problem solving, but also hamper its development.

It should be pointed out that most physicians engaged in a diagnostic work-up follow the same more or less standard procedure for eliciting data from a new patient. The usual sequence is: (1) history of present illness, (2) review of systems, (3) family and social history, (4) physical examination proceeding generally from head to feet, and (5) laboratory work and special tests. Such a standard sequence appears to be more consistent with routine data gathering than hypothesis testing. Barrows and Bennett (1972) noted, however, that the systematic, routinized search for data is made for the purpose of generating hypotheses. When seriously entertained hypotheses emerge from the data, Barrow's physicians temporarily abandoned the standard screening routine to pursue the hypotheses. Only after the physicians established the likelihood of their hypotheses at some satisfactory level, did they resume the routine screening procedure in a systematic search for further or refined hypotheses. During the course of a standard diagnostic work-up there may be numerous detours from the main sequence of data gathering, to the extent that the work-up may look superficially like a random art. The fact that most experienced physicians emerge from this combination of routine and seemingly random activity with similar diagnostic conclusions lends credence to the belief that although diagnostic behavior is less than standard, it is a well-ordered cognitive activity.

The observations and conceptualizations of Elstein and Shulman, Schwartz and Simon, and others are essentially in agreement with those of Kleinmuntz and Barrows and Bennett on the concept of hypothesis testing. Schwartz and Simon conclude that the data gathering of physicians is intended to: (a) definitely confirm hypotheses, (b) definitely reject hypotheses, (c) introduce more specific or refined hypotheses, (d) add to the likelihood of hypotheses, (e) lower the likelihood of hypotheses, or (f) add rival hypotheses to the pool.

Elstein has indicated in personal communication, however, that attempts by his research group to distinguish reliably hypothesis-testing inquiries from non-hypothesis-testing inquiries have been discouraging. Sets of decision rules for classification of physician inquiries on this dimension have been found to have little or no generality across specific diagnostic problems. The task of classifying inquiries as routine

or hypothes
purpose mus
question or
his purpos
one purpos
may be ent
infarctio
me more a
taneously
(b) esta
patient'
erate ne
chest wa

inquiry
collect
the dis
present

proble
signif
and se
simpl
emerg
three

or hypothesis testing is immensely difficult since the purpose must either be inferred from the context of the question or verified by asking the physician to recall his purpose later. Many questions may have more than one purpose. The physician inquiring about chest pain may be entertaining hypotheses such as myocardial infarction, angina, and pleurisy. The invitation, "Tell me more about your chest pain" may be intended to simultaneously: (a) differentiate between these hypotheses, (b) establish rapport with the patient, (c) determine the patient's emotional reaction to the pain, and/or (d) generate new hypotheses not previously considered, such as chest wall trauma.

The hypothetical constructs of hypothesis-testing inquiry, hypothesis-generating inquiry, and routine data collection may prove to be unproductive either because the distinction is in fact a false one, or because we presently lack the techniques to make the distinction.

In any case the research efforts in diagnostic problem solving begun barely a decade ago have already significantly increased our understanding of the process and seriously challenged the conventional wisdom. The simplified model of medical diagnosis that has so far emerged from research efforts might be summarized in three steps:

(1) Ir

pl

sa

i

(2) I

d

s

l

(3)

of medi

ing th

and th

ticed.

cessin

explai

highly

somewh

cally

tic b

(1) Immediate elicitation of the patient's chief complaint while scanning the patient for highly salient demographic characteristics and abnormalities of appearance or behavior;

(2) Initial grouping of cues and formulation of a few diagnostic hypotheses at varying levels of specificity based upon an extremely limited data base;

(3) Continuous development and evaluation of hypotheses by employing variations of a standard screening procedure for data elicitation. Data elicited serve the purposes of generating, modifying, and testing a consistently small group of diagnostic hypotheses.

Theoretical Constructs Underlying Diagnostic Problem Solving

One might justifiably ask why the logical advice of medical textbook authors and clinical faculty regarding the approach to diagnosis is so universally ignored and the hypothesis-guided approach so universally practiced. Two theoretical orientations, information processing theory and decision theory, seem appropriate to explain at least some of the observations. Both of these highly technical sets of theories will be presented in a somewhat simplified discussion before borrowing eclectically from them to arrive at explanations of the diagnostic behavior of physicians.

Information

The

can be trace

eighteenth

osophers ev

through the

Johnson, 1

was brough

Titchener

elusive ir

of subjec

condition

elude res

have rece

nical lar

tigation

theory c

Weaver

It was

Miller

by Mill

into ps

as an

enviro

inform

Information Processing Theory

The modern history of the psychology of thinking can be traced to the British Associationists of the eighteenth and nineteenth century. These logical philosophers evolved principles of the association of ideas through the analysis of their own mental experience (D. M. Johnson, 1972). The study of thought by introspection was brought into the twentieth century by Wundt and Titchener (Boring, 1950), who attempted to constrain elusive introspections by the use of more precise methods of subject training and experimentation under controlled conditions. The reasoning processes which continued to elude researchers during the early part of this century have recently been given a conceptual structure, a technical language, and a more precise methodology for investigation by information processing theory.

Information processing theory is essentially a theory of communication systems developed by Shannon and Weaver (1949) in the context of telephone engineering. It was introduced into cognitive psychology primarily by Miller (1956), Bruner, Goodnow, and Austin (1956), and by Miller, Galanter, and Pribram (1960). As translated into psychology, this theoretical orientation views man as an information channel receiving information from the environment and acting adaptively upon it. Between information input and behavioral or cognitive output,

the human in
by actively
manipulating
(Schwartz &

The
part respon
nitive psyc
a wide vari
problem sol
(DeGroot,
Reitman, 1
et al., 19

S
dations l
the chara
system wh
great var
mation:

> The
> proc
> elem
> mill
> proc
> with
> has
> memo
> that
> sec

and the
long-te

the human information processor operates upon the input by actively selecting, encoding, storing, transforming, manipulating, decoding, and retrieving the information (Schwartz & Simon, 1970).

The information processing paradigm has been in part responsible for the resurgence of research in cognitive psychology. It has been applied to analysis of a wide variety of cognitive tasks including general problem solving (Ernst & Newell, 1967), Chess problems (DeGroot, 1965), and concept formation (Hunt, 1962; Reitman, 1965) as well as medical diagnosis (Elstein et al., 1972; Wortman, 1972; Schwartz & Simon, 1973).

Simon and Newell (1972), building upon the foundations laid by Miller, Bruner, and others, have described the characteristics of the human information processing system which appear to be invariant over people and the great variety of tasks requiring rational use of information:

> The system operates essentially serially, one-process-at-a-time, not in parallel fashion. Its elementary processes take tens or hundreds of milliseconds. The inputs and outputs of these processes are held in a small short-term memory with a capacity of only a few symbols. The system has access to an essentially infinite long-term memory, but the time required to store a symbol in that memory is of the order of seconds or tens of seconds. [p. 149]

The small symbol capacity of short-term memory and the time requirements for transferring symbols into long-term memory are important constraints on the ability

of the hu
of inform
more effi
mentary "
individua
may come
ience. F
may posse
of letter
these ele
symbols r
His symbo
where on
numbers
to hold
the same
cessing
within t

informat
tainty a
that a d
achieved
The huma
it recog
convey m

of the human system to use information. Two principles of information processing help to make the human system more efficient. First, the symbols may represent elementary "bits" of information, or by transformation, individual bits may be "chunked" together and one symbol may come to represent larger and larger pieces of experience. For example, when first learning to read a child may possess symbols which represent only letters or parts of letters. As he develops reading skills he may "recode" these elementary bits of information successively into symbols representing whole words, phrases, and concepts. His symbol capacity remains essentially unchanged, but where once he was limited to holding a few letters or numbers in short-term memory, he may eventually be able to hold a few complex theories or mathematical proofs in the same space. A second principle of information processing derives from the definition of "information" within this system.

A particular datum carries varying quantities of information in proportion to its value in reducing uncertainty about the situation in question. To the extent that a datum is redundant with knowledge states already achieved it loses value as information (Attneave, 1959). The human system has the capacity to screen out data which it recognizes as redundant and to select those data which convey maximum information. In this way, we are able to conserve the limited time, space, and energy required to process the most essential aspects of experience.

Simon and Newell (1972) make the point that the space and time limitations inherent in human information processing still permit a wide range of flexibility in how problems are attacked. The empirical studies of problem-solving behavior in tasks ranging from simple laboratory manipulations to complex social and economic tasks demonstrate that there are almost as many theories of problem solving as there are tasks to be solved (Wason & Johnson-Laird, 1968). Simon and Newell attribute this variety to the flexibility of the system within its few constraints. The actual behavior in any given task situation is dependent not only on the few information processing constraints of the system, but largely on the nature of what Simon and Newell call the "task environment"--the characteristics of the task itself.

Returning again to the medical domain, the major characteristics of the task of diagnosis identified by the author of this study are as follows:

1. The task is taxonomic. The physician attempts to place the patient's problem in its proper place in a pre-established taxonomy of disease entities.

2. The taxonomy has a number of levels of specificity, with categories which may overlap, may be descriptive or etiologic, and may be incomplete.
3. The data which are useful in the taxonomy are selected by the problem solver in a sequential order determined largely by him.
4. Any datum selected by the problem solver is associated probabilistically with at least one taxonomic category. Data differ in reliability, diagnosticity, and in the cost of the interventions required to obtain the data.
5. The task of diagnosis presupposes that the problem solver is already well acquainted with a store of information about the structure of the taxonomy and its relation to the data from which the taxonomic state will be inferred.

Simon and Newell (1972) postulate that "subjects faced with problem solving tasks represent the (task) environment in internal memory as a space of possible situations to be searched in order to find that situation which corresponds to the solution." Translated into terms applicable to medical diagnosis we might say: Doctors faced with diagnostic problems represent the status of the environment in internal memory as a complex of historical data, signs, symptoms, laboratory values, and test results to be searched in order to find the diagnosis which corresponds to the actual pathology.

The size of any nontrivial problem space is enormous. Simon and Newell note that the problem space of a chess game is probably 10^{120} possible assignments of chess pieces to board position. The problem space of medical diagnosis is virtually indeterminate in size. Fortunately, Simon and Newell point out, the size of the problem space is not very important because we need not evaluate every possible combination of information states by trial and error. Instead, we employ heuristics--rules of thumb--which guide us to the examination of small, promising regions of the problem space where crucial data are most likely to be located.

Viewing the task of medical diagnosis as a problem which must conform to the constraints of the human information processing system and to the nature of the task environment, we can explain some of the observations of diagnostic performance reported by investigators of the process. Further, we can speculate on the effectiveness of different behaviors and strategies observed in practicing physicians. In a subsequent section of this review the concepts and principles of information processing theory will be applied in this way.

Decision Theory

A second way to view the diagnostic process is through the decisions that are made. Decision theories hold that choices are made under uncertainty about the eventual outcomes of the decision. But the experienced decision maker has at least some knowledge beforehand about the costs which his decision will commit him to and the probabilities of possible outcomes of his decision. The possible outcomes may be positive, negative, or inconsequential. The optimal decision maker will attempt to find a decision which balances all of the positive and negative aspects of the decision in a manner to produce the best expected outcome. Each of the aspects of the decision can be subjectively quantified in terms of the value or desirability of each outcome to the decision maker, the probability of each outcome under certain conditions, and the costs of the decision. The values, probabilities, and costs once quantified can be manipulated mathematically in order to yield what is called by some theoreticians a utility function. These theoreticians (von Neumann & Morgenstern, 1953; Edwards, 1954; Einhorn, 1971) have constructed theoretical models in which optimal decision making is defined by the choice which maximizes utility.

Although utility theory has generated much sophisticated mathematical work, it has to date had little impact on behavioral decision theory (Lee, 1971). For example, there is no generally agreed-upon function describing the utility of money among gamblers. Part of the problem in defining generalizable utility functions for humans in choice situations lies in the inherently unstable nature of utilities as defined by von Neumann and Morgenstern (1953). Utility in their sense incorporates subjective estimates of desirability and expected usefulness at the time of decision for a decision maker who may be either informed or ignorant, rational or irrational. Thus, for a child the utility of drinking iodine may be greater than the utility of drinking milk. As this example illustrates, utility functions may differ radically among individuals with different information, experience, levels of curiosity, or dispositions.

In medical diagnosis decisions are made continuously about what hypotheses to test, what data are worth gathering to test a given hypothesis, and when it is no longer worthwhile to continue choosing hypotheses or testing chosen ones. The choice of what initial hypothesis to test does not usually involve significant cost differentials, except that a decision to test a general hypothesis necessarily means that further, more specific hypotheses will need to be subsequently entertained. The second important factor in the choice is the probability that the testing of the hypothesis will bring the clinician significantly closer to the correct diagnosis. A negative consequence of choosing a particular hypothesis to investigate might be that while the chosen hypothesis is pursued, the patient's condition might deteriorate due to a nonhypothesized cause. The construct of value implies that different kinds of pathology are more important to the physician (and to the patient) than others. For example, obtaining a correct diagnosis of bronchial pneumonia would no doubt have greater value than obtaining an equally correct diagnosis of a common cold.

With respect to decisions about what data to gather in order to test a given hypothesis or set of hypotheses, the cost, probability, and value of the test may all be important factors in the decision. Costs may be as minimal as a minute or two of time if the test involves obtaining a smoking history. On the other hand some diagnostic laboratory procedures may involve large financial outlays, great physical and emotional discomfort, and even substantial risk to the patient's health. The probability of a diagnostic outcome of a test is a function of both the likelihood that the patient has the pathology the test is intended to assess and the reliability of the test itself. The value of a test is a function of its diagnosticity. For example, a finding of pathologic bacteria in the sputum is more diagnostic of pneumonia, i.e., has greater value, than a finding of low-grade fever.

In the following section the diagnostic behavior of physicians will be briefly re-examined employing the concepts of both information-processing and decision-making theory. These theoretical grounds are used to provide a partial justification for hypothesis-guided inquiry in general. In addition implications of the theories are used to rationalize various diagnostic processes and to suggest hypotheses about effective diagnostic reasoning.

Explanations of the Diagnostic Behavior of Physicians

The diagnostician when first confronted with a patient is faced with the task of selecting a small portion of the data from a potentially unlimited data pool. Each datum has greater or lesser importance given any of several hundred potential diagnoses. Several principles of information processing immediately come into play. First, the physician must restrict his search to currently available areas of the problem space with the greatest promise. He apparently does this by seeking information known from past experience to have high information value; chief complaint, age, sex, race, general appearance, etc. This information substantially changes the probabilities of various types of disease which the patient is likely to have. Still, in the first few minutes of the patient encounter, information is coming in at a rate which may far exceed both the space limitation of the physician's short-term memory and the time requirements of long-term storage. Consequently, he must select and recode the incoming information in a way which optimizes the use of his short-term memory storage capacity. Using a few highly salient problematic elements of the patient's case to index his hierarchically organized taxonomy of disease categories, the physician retrieves a few categories from some level in the hierarchy (they "pop" [Barrows & Bennett] into his head in a matter of milliseconds) which assume the character of diagnostic hypotheses. These hypotheses become the organizing principles for efficiently selecting the subsequent high-information data and for chunking it for representation in short-term memory. (Although the scenario is the writer's invention, the last sentence is in substantial agreement with the conceptualization of Elstein and Shulman, Schwartz and others).

Kleinmuntz (1968) has demonstrated that data obtained which do not fit one or more of the hypotheses currently entertained tend to be totally forgotten. This finding is consistent with the information processing constraints outlined by Simon and Newell. That is, a piece of data which is not seen as evidence for a currently viable hypothesis is not worth the time required to store it in isolation in long-term memory, and it is soon crowded out of the limited space of short-term memory by the continuous input of new data. Barrows and Bennett (1972) also report observing the phenomenon of forgetting data that are useful to the final diagnosis but irrelevant to hypotheses entertained at the time the data were elicited. Disturbing corroborative evidence of this phenomenon has been published by Williamson, Alexander, and Miller (1967). In a survey of hospital records they found that clearly abnormal values for routinely ordered urinalyses, hematocrits, and blood sugars were ignored in two-thirds of the cases in which these test results were inconsistent with the physician's diagnosis. A series of educational interventions failed to alter this pattern of apparent inability to process inconsistent information. Responses to these unexpected laboratory values were finally improved when laboratory personnel obscured the report data with flourescent tape, which the doctor had to physically remove in order to read the results.

It would seem that the quality of the hypotheses initially selected as the organizing principles for subsequent encoding of information would be crucial to the subsequent problem-solving performance. This is in fact the reason for the great research interest by the authors previously reviewed in the earliest stages of diagnostic problem solving. We might well ask what decision rules should govern the selection of the three to five hypotheses for which short-term memory space is available. Elstein et al. (1972) have found at least four considerations determining the order in which hypotheses are ranked by physicians for investigation.

(1) The statistical likelihood of the disease for patients in a particular population defined by age, sex, race, and other gross characteristics;

(2) The seriousness of the disease in terms of possible life threatening or incapacitating consequences;

(3) The treatability of the disease or the probable effectiveness of physician intervention;

(4) The novelty of the disease.

Simon and Newell conceptualize an effective problem-solving strategy as an attempt to search "highly promising regions of the problem space" or to find data that are most <u>likely</u> to help solve the problem. Clearly, hypotheses entertained on the basis of seriousness or treatability may be unlikely diagnoses, and the deliberate consideration of novel hypotheses appears to be a supremely inefficient search strategy.

Invoking the concepts borrowed from decision theory, however, it is clear that probability is only one factor in the utility function determining the choice of a hypothesis. Seriousness, treatability, and novelty enter the utility equation as values. Physicians would agree that the value of detecting a serious or treatable disease is greater than the value of detecting a trivial disease or one for which there was no effective treatment. It is more questionable whether the value associated with novelty can compensate for its low probability in the utility formula. Elstein et al. (1972) and Barrows and Bennett (1972) have defended the entertainment of novel hypotheses because they may keep the physician interested in the problem and because they may keep his mind open to unlikely but plausible alternatives. Although an interested physician is a valuable asset to problem solving, the interest is likely to be centered on the novel disease instead of its more likely alternatives, and interest may quickly disappear when initial tests of the novel hypothesis prove negative. Additionally, the function of keeping the physician's mind open might be as well served by a different, more likely alternative.

Returning to the considerations of seriousness and treatability we see that these considerations can usually be applied only to specific hypotheses. It can be agreed that bronchial pneumonia is more serious and possibly more treatable than the common cold, but it is probably meaningless to say that disease of the cardiovascular system is more serious or treatable than disease of the gastrointestinal system. Consideration of seriousness and treatability in ranking hypotheses for investigation necessarily restricts the scope of the early hypotheses when data justifying such restriction are seldom available. It would seem that more general early hypotheses would result in both greater thoroughness of search and greater efficiency of encoding and chunking of information. It may be that seriousness and treatability are considerations for Elstein's and Shulman's physicians because, like novelty, they increase the physician's interest in the case. It makes little difference in terms of patient care outcomes, however, whether a correct diagnosis of a serious and treatable disease such as diabetes is obtained or ruled out in the first 5 minutes of a diagnostic work-up or 15 minutes later. Thus, the writer is in agreement with Barrows and Bennett (1972) that, because general early hypotheses are more likely than early specific hypotheses to include the definitive diagnosis and permit the encoding and chunking of relevant information, they are probably more useful.

Two exceptions to the rule of early general hypotheses seem justified, however. First, some very common problems such as the common cold have such a high probability of occurrence given typical symptoms that such specific early hypotheses may be well justified. Second, emergent conditions requiring rapid and aggressive treatment such as myocardial infarction or acute appendicitis are associated with such large early detection values that even when probabilities are quite low the utility of investigating these specific possibilities at the earliest moment is very high.

Testing of Hypotheses

After the hypotheses have been generated they are tested by the acquisition of further data which are used as either confirming or disconfirming evidence. Again, it is the physician who determines which pieces of potential evidence are collected and in what order; and which pieces are judged to be not worth collecting. The physician also determines the way in which his evidence will be evaluated. What kind of rules should he follow in the accumulation and utilization of his evidence?

The information processing approach would dictate that the physician seek data which would maximally reduce the uncertainty of the diagnostic situation. Reduction of uncertainty as a strategy means that the problem solver would avoid the collection of evidence which is redundant with evidence already obtained. However, since the task environment of medical diagnosis is one in which data are probabilistically related to several diagnostic outcomes and vary with respect to both reliability and diagnosticity, apparent redundancies are frequently sought as confirmatory evidence. Under the information processing model, information is obtained until the problem solver judges that further data will have zero value in reducing the uncertainty of the diagnosis, either because the diagnosis has been solidly established or because all reasonable hypotheses have been ruled out.

In the decision-making approach, the choice of data to be gathered as evidence for hypothesis testing would be made on the basis of maximizing the utility of information. That is, the expected value of the information would be weighed against the cost of obtaining the information. As previously mentioned, cost in medical diagnosis can be viewed as some function of financial expense, discomfort to the patient, and risk to the patient's health resulting from the diagnostic procedure.

Operating in a decision-making mode, the physician would first seek information of high expected value and low cost in order to maximize the utility of his evidence. This is one reason why the standard diagnostic work-up proceeds from history, to physical examination, to laboratory work. The diagnostic activity should continue until the expected value of the evidence reaches a zero point because further evidence would not alter the diagnosis or until the physician's subjective likelihood estimate of the cost of further evidence outweighs the expected value of the diagnostic evidence. At this point the information has a negative utility and evidence gathering should cease. Illustrations of this kind of reasoning are commonly seen in ambulatory care settings and in emergency medicine. For example, once a bacterial cause is established for a case of otitis media, the expected value of the precise identification of the pathogen is often seen as being outweighed by the expense and inconvenience of obtaining a culture, since broad-spectrum antibiotics are available which have a high probability of eradicating the infection regardless of the specific pathogen. In the emergency care setting, the physician who recognizes a case of shock will usually not immediately attempt to diagnose the etiology since the risk of delayed action outweighs the expected value of a more specific diagnosis.

Errors in Diagnostic Reasoning

The studies of the diagnostic process so far reviewed have concentrated primarily on the skills of practicing physicians and on optimal strategies derived from information processing and decision-making theories. The interest of the present study is in training of advanced medical students for the purpose of reducing the errors of reasoning commonly encountered in the hypothesis-guided approach. An informal survey of clinical faculty at Michigan State University has revealed the following judgments of common diagnostic problem-solving errors among students:

(1) Failure to plan and organize an approach to the problem;

(2) Failure to recognize more than one diagnostic possibility at a time;

(3) Failure to adequately synthesize the data on hand;

(4) Jumping to conclusions prematurely;

(5) Failure to let go of a disproven hypothesis;

(6) Over-dependence on ostensibly pathognomonic[1] signs;

(7) Ordering too many laboratory tests;

(8) Failure to recognize one's own limitations.

Adding to these anecdotal observations, Barrows and Bennett (1972) have systematically compared the

---

[1] A pathognomonic sign, in its common usage, is one which is uniquely associated with a particular disease. The finding of a single pathognomonic sign would provide a clinician with sufficient evidence to identify the disease with virtual certainty. With increased understanding of clinical signs and their relationships to underlying pathology, the usefulness of the concept has been called into question.

diagnostic process of medical students, residents, and practicing physicians. They reported that the diagnostic hypotheses of students tend to be unjustifiably specific, probably contributing to many students' inability to synthesize findings and to increased forgetting of data elicited. The errors mentioned by the clinical faculty members and Barrows and Bennett are probably most prevalent among students but can also be found in the performance of more experienced clinicians. Analysis of the problem-solving protocols of Elstein and Shulman's physician subjects confirms several of the clinical faculty's anecdotal observations and adds to the list of errors the failure to use negative information to disconfirm hypotheses.

Some of the errors described will have a familiar ring to those acquainted with the general literature on thinking and reasoning. In the following subsections, the errors will be classified and discussed with respect to theoretical positions and empirical support.

A. Failure of Planning

Miller, Galanter, and Pribram (1960) devoted considerable space to the importance of planning an approach to a problem. In their chapter on "Plans for Searching and Solving" they provided the example of a homeowner's search for a hammer. Such a search may begin with a planless wandering from room to room; it may be algorithmic with each room being systematically and exhaustively searched; or a heuristic plan may be devised to search first those places where the hammer is most likely to be found. The planless search certainly requires the least cognitive and physical effort, but unless the hammer is in plain view, the search is unlikely to be productive. The exhaustive algorithmic search is almost certain to turn up the hammer eventually, although with the prospect of a lengthy search a better alternative might be to borrow or buy another one. Physicians have a much less clearly defined area in which to search for diagnoses, and patients with obscure problems cannot be so easily dismissed as lost hammers. The third alternative--devising a heuristic plan--is the only feasible approach for searching for nontrivial solutions to problems located in very large spaces. Although this conclusion may seem obvious, the formation of high-quality plans is not universal or automatic.

In his treatment of heuristics Polya (1957) has distinguished four steps in the heuristic process. Of these steps the second--devising a plan that will guide the solution and connect the data to the unknown--is considered by Miller, Galanter, and Pribram (1960) as well as by Polya to be the most critical and the most creative. Polya's statement below underscores the importance, the difficulty, and the creative nature of this step:

> We have a plan when we know, at least know in outline, which calculations, computations or constructions we have to perform in order to obtain the unknown. The way from understanding the problem to conceiving a plan may be long and tortuous. In fact, the main achievement in the solution of a problem is to conceive the idea of a plan. [p. 8]

The medical student faced with a complex diagnostic task may have to devise and revise a series of plans and subplans required at different stages of his inquiry. The effort is demanding in the face of his own expectations and the expectations of most patients that he act promptly and decisively.

B. Failures of Hypothesis Specificity

Barrows and Bennett (1972) as previously mentioned have found hypotheses of students and residents to be unjustifiably specific in the early stages of a work-up. Corroborating data on this point have not been sought by other investigators. Allal (in progress) is presently engaged in a study in which she will attempt to characterize the structure of the set of problem formulations that emerge during the initial portion of a diagnostic work-up. The study will analyze the initial problem formulations of practicing physicians and train second-year medical students to be able to structure their formulations in ways similar to experienced clinicians.

C. Failures to Devise Competing Hypotheses

The errors of failing to recognize more than one diagnostic possibility at a time, failing to adequately interpret the data on hand, and jumping to conclusions seem to go hand in hand. The reason for the phenomenon was described in a classic article by Chamberlin in 1890 and reprinted most recently in Science (1965).

> The moment one has offered an original explanation for a phenomenon which seems satisfactory, that moment affection for his intellectual child springs into existence . . . and it grows more and more dear to him. . . . So soon as this parental affection takes possession of the mind, there is an unconscious selection and magnifying of the phenomena that fall into harmony with the theory and support it, and an unconscious neglect of those that fail of coincidence. . . . There springs up also an unconscious pressing of the theory to fit the facts, and a pressing of the facts to make them fit the theory. When these biasing tendencies set in . . . the search for facts, the observation of phenomena, and their interpretation, are all dominated by affection for the favored theory until it appears to its author . . . to have been overwhelmingly established.

Chamberlin offers a way to avoid the "partiality of paternalism" through his Method of Multiple Working Hypotheses. Under this approach, the scientific problem solver gives birth to a family of tentative hypotheses which can be weighed against each other in a more impartial manner. More importantly, the various hypotheses suggest different lines of inquiry that might otherwise be neglected. The problem solver is no longer obtaining evidence to support his single hypothesis, but instead to distinguish between his several hypotheses. The outcome of such a search according to Chamberlin is more likely to be a complex explanation. In a passage very appropriate to medical diagnosis, Chamberlin states, "We are so prone to attribute a phenomenon to a single cause, that, when we find an agent present, we are liable to rest satisfied therewith, and fail to recognize that it is but one factor, and perchance a minor factor, in the accomplishment of the total result."

Covington, Crutchfield, Davies, and Olton (1972) have recently published a well-researched program for facilitation of problem-solving skills among school children. A major component of the program is the devising of multiple working hypotheses, and the results obtained lend support to Chamberlin's method.

D. Failures to Re-interpret Data

Often in the course of a diagnostic work-up, an unexpected finding will give birth to a new hypothesis. When this occurs the newest brain child may be subject to the recency effects explored in great detail by verbal learning researchers. In any case, new hypotheses seem to hold new hope for a diagnostic solution and are pursued enthusiastically--especially if preceded by confusion. As previously mentioned, however (Kleinmuntz, 1968), data elicited previous to the formulation of a new hypothesis are not automatically associated with it, and a conscious effort is often required to reintegrate previous findings with respect to the new hypothesis.

E. Failures of Negative Inference

In his popular book, How Children Fail, John Holt (1964) described a scene in which children were asked to find a number between 0 and 10,000. The children asked whether the number was less than 5000. On being told no, their response was clearly one of disappointment. Children at this stage of intellectual development did not recognize that the negative reply to the inquiry was every bit as diagnostic as a positive reply would have been. While we may be charmed by such naivete in children, Elstein and Shulman have observed that their physician subjects did not take full advantage of negative inference. The reluctance or inability of humans in general to search for and extract information out of negative findings has been demonstrated in a series of simple but elegant experiments by Wason (1968). In one such study the subjects were told that the series of numbers 2, 4, 6 conformed to a rule which they were to find by generating series of their own. Following each series the experimenter told the subject whether the series conformed to the rule. When a subject was quite certain that he had discovered the rule, he was to announce it. Subjects of this experiment (Harvard students) followed a strategy of generating series which were positive instances of the rule they had hypothesized and seeking positive confirmation. A typical set of generated series was 8, 10, 12; 14, 16, 18; 20, 22, 24; 1, 3, 5. The subject inferred from the hypothesis tests that the rule was, "starting with any number, two is added each time to form the next number." In fact, the rule was "any three numbers in ascending order." Subjects failed to recognize that the only way to be sure of the rule was not by repeated confirmations but by violating the hypothesized rule and receiving disconfirmations. The subject-generated series illustrated above conforms to many possible rules including the subject's erroneous rule, the correct rule, and rules such as any whole numbers, any numbers greater than zero, any numbers less than 30, etc. None of these hypotheses were disconfirmed by the subject's hypothesis tests.

In his 1964 article on "Strong Inference," Platt advanced the argument that some fields of science achieve more rapid advances than other fields because they design experiments which will exclude at least one hypothesis. The more slowly moving sciences and scientists are bound to one hypothesis or one method which fails to exclude alternatives. Platt provided a simple test of the usefulness of a problem-solving inquiry: given any hypothesis the question should be asked "What [test] could disprove your hypothesis?" or given any test, "What hypothesis does your [test] disprove?" Platt's method of Strong Inference, while seeming counter-intuitive or at least mildly uncomfortable for most of us appears to be highly efficient and to be well rooted in the philosophy of science.

The Effects and Effectiveness of Problem-Solving Heuristics

In the preceding section of this review the common reasoning errors of medical students and physicians were reviewed and strategies of approach or heuristics were outlined which, if followed, are believed to be helpful in avoiding or minimizing their errors. At this point it is appropriate to review attempts to teach people to use heuristics to improve their problem-solving performance. The question of interest is whether individuals can learn to use their knowledge more productively in solving problems by adhering to a set of heuristics.

Recent interest in heuristics stems principally from Polya's popular book, How to Solve It, first published in 1945 and republished in an expanded second edition in 1957. Polya has expressed his belief that, at least in mathematics, knowledge of the process of problem solving is more important than knowledge of the content of mathematics:

> Our knowledge about any subject consists of "information" and "know-how." In mathematics "know-how" is the ability to solve problems and it is much more important than mere possession of information. You have to show your students how to solve problems. . . . (Polya, 1958, p. 102)

The question of content versus process is one that is sometimes hotly debated in American medical schools, but it is safe to say that some knowledge of both is indispensable for medical practice.

Polya approaches the teaching of problem solving by asking students questions of a particular kind. The questions are not intended to be hints in the solution of a particular problem, nor do they fit the model of a Socratic dialog. Instead, Polya's heuristic questions have two required characteristics, common sense and generality. In Polya's words:

> As they proceed from plain common sense they very often come naturally; they could have occurred to the student himself. As they are general, they help unobtrusively; they just indicate general direction and leave plenty for the student to do. (1957, p. 4)

Some examples of Polya's heuristic questions are as follows: What is the unknown? What are the data? Do you have a related problem? Could you solve a part of the problem? Polya's expectation is that by repeatedly asking these and similar questions of a student, the student will become aware of fruitful problem-solving approaches and will begin to ask these questions of himself independently.

A few empirical studies of the effectiveness of Polya's heuristic method have been completed. Larsen (1960) conducted an experiment using three sections of an introductory college calculus class. One section was taught by him using Polya's heuristic approach, a second section was taught by him using a conventional approach and a third section was taught in a conventional mode by a more experienced colleague whom Larsen considered to be a superior teacher. Three dependent measures were used to compare course performance in the three classes; a portion of the final examination emphasizing content and not expected to be influenced by the heuristic approach, a portion of the final examination using word problems in which the heuristics taught were expected to provide some help to the experimental group, and a special test designed specifically to assess student ability to use the heuristics taught to the experimental class.

The results of Larsen's experiment were equivocal. On the content items of the final exam the class taught by Larsen's colleague outperformed both Larsen's heuristics class and his conventional class, with the heuristics class performing worst of all. On the word-problem items of the final exam, Larsen's colleague's class decidedly outperformed Larsen's students and there was no significant difference between Larsen's heuristic and conventional classes. On the specially prepared heuristic test Larsen's heuristic section performed best, followed by the conventional class taught by the colleague, followed by Larsen's conventional class. All differences were significant at least the .05 level of significance.

Larsen drew the following inferences from his results:

> (1) A skilled teacher, using a "conventional" approach can help his students to achieve a significant mastery of routine calculus problems, without any significant sacrifice of ability to handle the kinds of problems appearing on the heuristic tests used in this experiment.
>
> (2) There is some indication that a heuristic emphasis in teaching elementary calculus can help students learn to handle the kinds of problem appearing on the heuristic tests used in this experiment.

In a related experiment Larsen (1960) presented three calculus problems to two groups of students. Following each problem, students were given a written debriefing on the correct solution of the problem and the experimental group was additionally given a single heuristic suggestion which would be helpful in solving the following problems. Scores on the three problems were equivalent between groups, but the group given the heuristic suggestion was able to significantly reduce the time required to obtain the solution to the third problem. It seems reasonable to conclude that the experimental group was able to save steps by using the heuristic suggestion, but the fact that only one heuristic was provided which had direct applicability to all problems raises the question of whether the suggestion was a general heuristic in Polya's sense or simply a broad hint.

In 1962 Ashton employed a more adequate design to compare the performance of ninth-grade algebra students under a heuristic and a textbook approach to the subject. One algebra teacher from each of five schools was selected. Each teacher taught two sections of ninth-grade algebra--one section by the heuristic method and one section by the textbook method. Pretest-posttest gain scores showed significantly greater achievement for each of the heuristic groups over its textbook course cohort. Unfortunately, no analysis was made using intact classes as the unit of analysis. Ashton's results are encouraging evidence of Polya's position but leave unclear just how the employment of heuristics might influence problem solving. Because the teaching of the heuristics in this study was inseparable from the teaching of the course content, the use of heuristics may have influenced the amount of content learned, may have increased the students' ability to effectively apply their knowledge of mathematical content to the test problems, or both. The previously cited Larsen study did attempt to get independent measures of the content and process effects of heuristics, but design limitations preclude meaningful analyses of his results with respect to this question.

Wilson (1967) conducted a series of experiments investigating the effects of heuristic training on problem solving using mathematical material that had been previously learned and using unfamiliar material involving symbolic logic. Specifically, Wilson varied the level of generality of the heuristics taught and the order of presentation of the familiar and unfamiliar material. His particular results are too complex for discussion here, but the overall interpretation was that heuristics, either general or specific, facilitate the effective use of previously learned material, that the combination of specific and general heuristics may be complementary, and that greater positive transfer in the use of heuristics to dissimilar material can be expected from more general heuristics. Interactions between level of specificity of the heuristics and familiarity with the material led to the conclusion that in the presentation of new material, specific heuristics should precede more general heuristics. With familiar material, specific heuristics are of more limited value. Wilson's conclusions appeared to be adequately tested, using appropriate design and analysis techniques.

Outside of the domain of mathematics only a few studies are available in which heuristics have been manipulated as an independent variable. Perhaps the earliest was a well-known experiment reported by Maier in 1942. As part of a series of investigations on productive thinking, Maier had two groups of subjects solve an insight problem. The experimental group was first given a brief lecture including problem-solving hints which would probably qualify as heuristics in Polya's sense of the term. The hints or heuristics were as follows:

> 1. Locate a difficulty and try to overcome it. If you fail, get it completely out of your mind and seek an entirely different difficulty.
> 2. Do not be a creature of habit and stay in a rut. Keep your mind open for new meanings.
> 3. The solution pattern appears suddenly. You cannot force it. Keep your mind open for new combinations and do not waste time on unsuccessful attempts. [p. 147]

The three hints are largely redundant, and the kernel of the advice--keep your mind open and moving--is quite general. The group of subjects receiving this advice performed significantly better than the control group. The experiment was repeated with similar results.

Maier attributed the superior performance of the experimental group to the addition of "direction" to his subjects' necessary but not sufficient fund of past experiences. A question of interest is whether the particular direction provided by general heuristics is important, or is any direction sufficient. In other words, do heuristics facilitate problem solving only when they point out the right direction, or do they produce a heightened attention to the problem and a greater generalized alertness to whatever process the problem solver may be employing. Maier's experiment might be repeated with the addition of a third group given instructions diametrically opposed to his original hints to test this possibility.

Loupe (1969) conducted a study in which college students were given heuristic training to improve their problem-solving skills. The training materials consisted primarily of modified Sherlock Holmes mysteries which the subjects solved by requesting and obtaining information about the case which they believed might be helpful.

After each new piece of information received, the instructor asked the following six heuristic questions of the experimental group subjects:

1. Were there any new problems in the information?
2. What were the important details presented?
3. How did the new findings relate to the problem definition (problem redefinition)?
4. Was the hypothesis under test confirmed or disconfirmed?
5. Were there any new hypotheses?
6. Which of the possible hypotheses should you test, and where would you expect to find relevant information?

In the posttest phase of the study similar mysteries were solved by the students without the aid of the heuristic questions. Loupe reported that on his measure of problem-solving quality, posttest scores of the experimental group were significantly higher than those of the control group ($p < .025$).

The training effect of improved problem-solving quality did not transfer to problem-solving performance on another task of entirely different content and format. The transfer test--The Teacher's In-Basket (Shulman, Loupe, & Piper, 1968)--required the subjects to assume the role of a substitute teacher and to identify and solve student problems through the use of information found in the teacher's in-basket. These materials included the cumulative records of each child in the class as well as various potential information sources such as telephone messages, schedules, and memos.

Loupe's results suggest that training in problem-solving heuristics in nonmathematical domains can be effective in improving problem-solving quality, but that transfer to problems of dissimilar content is not to be expected.

The empirical studies reviewed provide some evidence that heuristic training methods have resulted in improved problem solving. The body of evidence supporting the effectiveness of heuristics is small, however, and there are several qualifications which limit the generalizability of the conclusions. Larsen's results serve as a reminder that good teaching requires first the ability to communicate an adequate understanding of the subject matter, and that given an adequate understanding of the material students may well be able to provide useful problem-solving strategies of their own. Wilson's study suggests that the effectiveness of heuristics may depend upon a match between the familiarity of the student with the content and the generality of the heuristics taught. Loupe's results suggest that general heuristics learned in the context of a particular type of problem cannot be expected to generalize automatically to other kinds of problems even when the same heuristics are applicable. The research of both Larsen and Maier raise the question of the distinction between heuristics, which may sometimes lead the problem solving astray, and hints, which are intended to guide the problem solver always in the correct direction. Finally, the small body of research on heuristics leaves unanswered the question of whether heuristic training influences only the ability to manipulate learned subject matter or influences the actual amount, type, or structure of the learned subject matter as well.

CHAPTER III

PROCEDURES

Subjects

The subjects of the study were medical students who had recently completed three years of undergraduate medical education and were beginning their fourth year of training at either Michigan State University or the University of Michigan. Students were contacted by telephone and asked to participate at a rate of $5.00 per hour as subjects in a study of diagnostic problem solving. Subjects were contacted in random order and randomly assigned to one of four treatment groups. Letters of confirmation were sent to each subject (see Appendix A). Only one student in each of the two medical schools declined to participate. All other students contacted either agreed to participate or were unable to take part due to an inability to find a mutually convenient time to schedule experimental sessions.

Development of Diagnostic Cases

Each of the four medical cases developed for this study had its origins in an actual case history.

Actual findings were extensively modified, however, to eliminate some misleading cues and to add dimensions to make them appropriately challenging to the population of interest. The original data base was extensively elaborated upon in an attempt to anticipate any reasonable piece of information that might be requested in a complete medical history, physical examination, or laboratory search for diagnostic cues.

In order to ascertain the adequacy of the simulated data base the cases were pilot tested by physician and student subjects. In addition, physicians were asked to review the entire data base of each case in order to detect conflicting information or conspicuous omissions. A final list of positive findings and important negative findings was compiled for each case (see Appendix C).

The cases were judged by the physicians to represent nontrivial diagnostic problems in the discipline of Internal Medicine which could be solved with near-certainty given the available data. The physicians expected the cases to be challenging but within the range of ability of competent fourth-year medical students. Each case was designed to permit elicitation of important diagnostic information in the medical history, physical examination, and laboratory stages of the work-up. The correct diagnoses of cases 1, 3, and 4 were structured to include one highly significant primary medical problem, three to four complications or manifestations of the primary problem, and one or two unrelated minor problems. The correct diagnosis of case 2 had a slightly different structure since two highly significant medical problems were present. Each of the four cases is described in the following paragraphs.

Case 1--Man with Complete Exhaustion: In case 1 a male college student presents with complete exhaustion. Further historical probes lead to a clinical picture of gastrointestinal distress including abdominal pain, weight loss, and abnormal bowel function. Findings on physical examination include severe anemia, generalized weakness, and abdominal tenderness. Laboratory results and x-rays localize the problem to the lower gastrointestinal tract and confirm a diagnosis of ulcerative colitis.

Case 2--Woman with Fatigue and Headache: In case 2 a female college student complains of headache, fatigue, and fever arising in the past week. Historical data are suggestive of an acute infectious process, probably of viral origin. Physical examination reveals an injected pharynx, pallor, and an enlarged spleen. Laboratory tests confirm a diagnosis of infectious mononucleosis and indicate the presence of an anemia. Further tests uncover a more significant problem of hereditory spherocytosis which has been dormant until apparently triggered by the acute episode of mononucleosis.

Case 3--Man with Left Chest Pain: Case 3 involves a 40-year-old construction worker complaining of left chest pain. The pain initially appears to be of either cardiovascular or of pleuritic origin. Further historical inquiry reveals progressive fatigue, poor appetite, significant weight loss, and skeletal pain at other sites. Physical examination reveals pallor, an enlarged spleen, and localized tenderness at specific rib, spine, and skull locations. Important laboratory findings include anemia, abnormal blood-forming cells in the bone marrow, multiple skeletal lesions including a pathologic rib fracture and abnormal proteins in the urine. The entire clinical picture is consistent with the disease multiple myeloma, a malignant plasma cell dyscrasia.

Case 4--Woman with Nausea and Vomiting: The patient in case 4 is a young mother whose chief complaint is nausea and vomiting. Her illness began less than a week prior to the clinic visit, progressing from a slight malaise to her present chief complaint as well as throbbing headache, dizziness, loss of appetite, shortness of breath, and heart palpitations. On physical examination she has a markedly elevated blood pressure, heart rate, and respiratory rate; flank tenderness, pallor, lung congestion, heart murmur, and ankle edema. The clinical findings are suggestive of congestive heart failure, but lack of evidence of previous cardiovascular problems leads to considerations outside the cardiovascular system. Laboratory results such as a routine urinalysis, blood counts, and other specific tests confirm a diagnosis of acute glomerulonephritis, an immunologic reaction affecting the capillaries in the kidneys. The disease is directly responsible for most of the symptoms, including the complication of congestive heart failure.

Case 1 was used as a pre-test for all subjects; case 2 was used as a training case for those subjects receiving training in the experimental heuristics. Cases 3 and 4 were used as posttest cases for all subjects. Subjects receiving training in the experimental heuristics were also given a debriefing on their pre-test performance with respect to their use of heuristics.

Format and Presentation of the Diagnostic Cases

The manner of presentation of the diagnostic problems was the same for all cases and all subjects. The verbatim instructions to subjects and the case data comprising the cases are presented in Appendices A and B.

An appointment was arranged in which the experimenter met with each of the 32 subjects on an individual basis for an average of four hours. Each experimental session began with the reading of the "Instructions for all Subjects" (see Appendix A). The subject was then given a single sheet of paper with a brief paragraph describing the clinical setting, the patient's general appearance in gross terms, and the circumstances under which the patient arrived at the outpatient department of the hospital. In addition, a brief list of routinely gathered data was provided including occupation, height, weight, age, temperature, and chief complaint.

The task of the subject was to ask for whatever additional information he desired including medical, social, and family history, physical examination results, and any laboratory or instrumental procedures in order to reach a diagnosis of the case.

Each five consecutive pieces of information elicited by the subject constituted a numbered search phase. At the completion of each search phase, the experimenter interrupted the subject to ask him what problem formulations or hypotheses he was presently considering. The subject continued to request and receive information until he was satisfied with his diagnosis, or until he decided to "refer" the case. At this point the subject was asked to complete a Diagnostic Summary Form (see Figure 3) which outlined his diagnosis.

Independent Variables

The independent variable of greatest interest was the type of heuristic training and prompting to which subjects were exposed. Subjects were randomly assigned to one of four treatment groups as described in the following paragraphs.

Treatment Group 1. Subjects in Treatment Group 1 were read the initial instructions for all subjects and proceeded to solve the pretest problem, case 1. Upon finishing case 1, subjects were given an introduction to the use of heuristics and an explanation of each of the five experimental heuristics along with the rationale for its use and several examples of where and how they might be used. When the subject indicated that he understood the use of the heuristics, he was debriefed on his pretest performance with emphasis on how the use of the experimental heuristics could have improved his performance. In the next part of the training, case 2 was presented to the subject and at least one of the five heuristic questions was asked of the subject after each five pieces of information elicited. The particular question asked was dependent upon the current status of the problem. The experimenter actively helped the subject in interpreting the heuristic question, suggesting alternate replies that might alter the course of the problem-solving effort, and in general facilitated the subject's effective use of the heuristic questions. No performance data were collected on the training case. Following the training phase, subjects were presented with two posttest problems; cases 3 and 4. After each group of five pieces of information elicited subjects were asked to review a printed list of the heuristic questions, to select at least one question which was appropriate to his current status in the problem, and to verbalize an answer to that question.

Treatment Group 2. Subjects assigned to group 2 received instructions and treatment identical to group 1 with one exception. In the posttest phase, the pretest list of heuristic questions was placed before them with instructions to ask themselves the heuristic questions as they solved the problems. That instruction was given once before the presentation of each of the two posttest cases.

Treatment Group 3. As in groups 1 and 2, subjects in group 3 were read the initial instructions to all subjects and proceeded to solve case 1, the pretest case. At the completion of case 1, subjects were told the correct diagnosis, were given some nonheuristic feedback on their performance, and were permitted to ask questions about the case. Following this discussion, each subject was given an orientation into the use of heuristics. Each subject was asked to generate from his own experience four to six rules of thumb which had been helpful to him personally as guides to diagnostic problem solving.

Subjects were then presented with the posttest problems. Following each five pieces of information elicited, subjects were asked to review their idiosyncratic heuristics, to select at least one which was appropriate to the current status of the problem, and to verbalize an answer to that question.

Treatment Group 4. Subjects in group 4 were presented with the pretest case under the same conditions as all other groups. Discussion of the case immediately afterwards was permitted. Subjects were then given a brief orientation in the use of heuristic rules of thumb and were asked to be aware of their own rules of thumb as they attempted to solve the remaining problems. Subjects were then presented with the two posttest problems, cases 3 and 4.

The second independent variable of the study was the medical school of enrollment. Subjects were selected from two Michigan medical schools. One school was a well-established, highly respected medical college; the other was a new medical school with what was generally considered to be a more innovative and more flexible curriculum. Subjects from the two schools were comparable in their Medical College Admission Test Scores at entry (School A MCAT mean = 574, S.D. = 67; School B MCAT mean = 556, S.D. = 56; $t = .91$; $p. > .25$). The two schools did not appear to differ substantially in their approaches to diagnostic training, and no differences between schools on any of the dependent measures were expected. The principle reason for the inclusion of the two medical schools in the sample was the extension of the external validity (Campbell & Stanley, 1963) of the results. The plan of the experiment is represented schematically in Figure 1.

Construction and Analysis of Dependent Measures

Ideally, the psychometric properties of dependent measures should be well known prior to their incorporation in experimental studies. Unfortunately, the adequate prior testing of the dependent measures of this study would have required an additional 100 hours of individual testing and would have exhausted the available subject pool for a period of one year. Consequently, the subjects of the study served two purposes. In addition to testing the principle research questions, they also provided an empirical data base required to test the reliability of newly created dependent measures. The following sections describe each of the variables and report the results of appropriate reliability studies.

Four measures of performance were obtained from each subject on each diagnostic case presented. The first measure was intended to capture the range or scope of the subject's diagnostic formulations based on the

	Treatment Group	Pretest	Training	Posttest
Medical School 1	T_1	Pretest (Case 1)	a. Exper. Heur. Orient. b. Pretest Critique c. Training (Case 2)	Posttest (Cases 3 & 4) with Systematic Exper. Heuristic Prompting
	T_2	Same as T_1	Same as T_1	Posttest (Cases 3 & 4) with no Heuristic Prompting
	T_3	Same as T_1	a. Gen. Heur. Orient. b. Ident. of Idiosyn. Heuristics	Posttest (Cases 3 & 4) with Systematic Idiosyn Heuristic Prompting
	T_4	Same as T_1	a. Gen. Heur. Orient.	Posttest (Cases 3 & 4) with no Heuristic Prompting
Medical School 2	T_1	Same as T_1	a. Exper. Heur. Orient. b. Pretest Critique c. Training (Case 2)	Posttest (Cases 3 & 4) with systematic Exper. Heuristic Prompting
	T_2	Same as T_1	Same as T_1	Posttest (Cases 3 & 4) with no Heuristic Prompting
	T_3	Same as T_1	a. Gen. Heur. Orient. b. Ident. of Idiosyn. Heuristics	Posttest (Cases 3 & 4) with Systematic Idiosyn Prompting
	T_4	Same as T_1	a. Gen. Heur. Orient.	Posttest (Cases 3 & 4) with no Heuristic Prompting

Fig. 1. Plan of Experimental Procedures

data elicited in the early phases of a diagnostic work-up. The scope of the early hypotheses is a concern of the hypothesis-guided approach because it sets limits on the kinds of information judged to be most useful in the subsequent search for data. Therefore, it is of interest to determine how the scope of early diagnostic formulations will vary among medical students trained in various modes of heuristic problem solving and how these variations might influence diagnostic outcomes.

The second measure was intended to determine whether different conditions of training and usage of heuristics would influence the number and importance of diagnostic findings elicited by the diagnostic problem solver. The hypothesis-guided method places emphasis on efficiency rather than thoroughness, but it was not clear whether greater efficiency of information search would reduce the number of critical findings elicited or whether a reduction in critical findings elicited would significantly influence the outcome of the problem-solving effort.

The two measures described above may be considered as process measures since they are indices of activities performed in the process of a diagnostic work-up.

The two remaining measures--cost of the work-up and accuracy of the definitive diagnosis--can be considered as outcome measures central to the evaluation of diagnostic competence. More detailed description of the four dependent measures, their rationale, and research hypotheses are given below.

Scope of Early Diagnostic Formulations

Analysis of the protocols of physician performance in the studies of Elstein and Shulman revealed that although several hypotheses may be simultaneously entertained, they may be related to each other in various ways. First, they may be multiple competing hypotheses in the sense of the term used by Chamberlin. For example, a physician might say, "This patient's headache is probably due to tension, but I'd like to make sure that there isn't an organic cause instead." Second, hypotheses may be functionally related. For example, a physician might say, "The fracture could be due strictly to trauma, but I suspect that an underlying disease might have weakened the bone, predisposing the patient to fractures." Third, a group of hypotheses may be hierarchically arranged with the more specific hypotheses presented as exemplars of a broader class of disease. An example of this kind of structure is contained in the following hypothetical statement: "This looks like an acute bacterial infection such as bronchitis or pneumonia." Further, the competing, functionally related, and hierarchical structures may be combined. Consider the following formulation in which all three structures are included: "This patient's chronic fatigue and anemia could be due solely to his poor dietary habits or perhaps to some kind of cancer affecting erythropoisis such as multiple myeloma or Hodgkin's disease. Of course, diseases like this often decrease the appetite so we may have a vicious circle."

Analyses of isolated hypotheses to determine their number or specificity are bound to gloss over the structural relationships between them. What is of greatest interest to researchers operating within an information-processing paradigm is the proportion of the hierarchical taxonomy of diseases included in the hypotheses under which case data will be subsumed.

Subjects were asked to generate hypotheses after each five pieces of information elicited. The first four sets of hypotheses generated (through the 20th piece of information elicited) constituted the total set of early diagnostic formulations. The measure, Scope of the Early Diagnostic Formulations, was defined by displaying subjects' early hypotheses as an area within a multi-level disease process by organ system grid representing the total space of the disease taxonomy. The area of the grid covered by a subject's hypotheses constituted his score on this measure.

Since the area covered by one hypothesis may partially overlap with, or be completely contained within another hypothesis, the Scope score automatically takes into consideration the relationship between hypotheses; giving additional credit for each new hypothesis only to the extent that the new hypothesis opens areas for investigation in the disease taxonomy not included under previously considered diagnostic formulations. Figure 2 illustrates the disease process by organ system grid scored for a hypothetical subject having four hypotheses.

Particular diagnostic formulations by subjects may be a function of the specific data currently on hand for a specific problem as well as knowledge of medicine, clinical experience, and problem-solving style. Consequently, the stability of the Scope measure within problems as more data accrues or across problems of different content are not meaningful considerations in establishing the reliability of the measure. Inter-rater reliability is of importance, however. Two trained raters independently scored the Scope of the Early Diagnostic Formulations for each of the three test problems. Pearson product moment correlations between the scores obtained by the raters are reported in Table 1 as evidence of the inter-rater reliability of the measure. Reliability estimates of average ratings were not computed since the dependent measures were based on the scoring by only one of the raters.

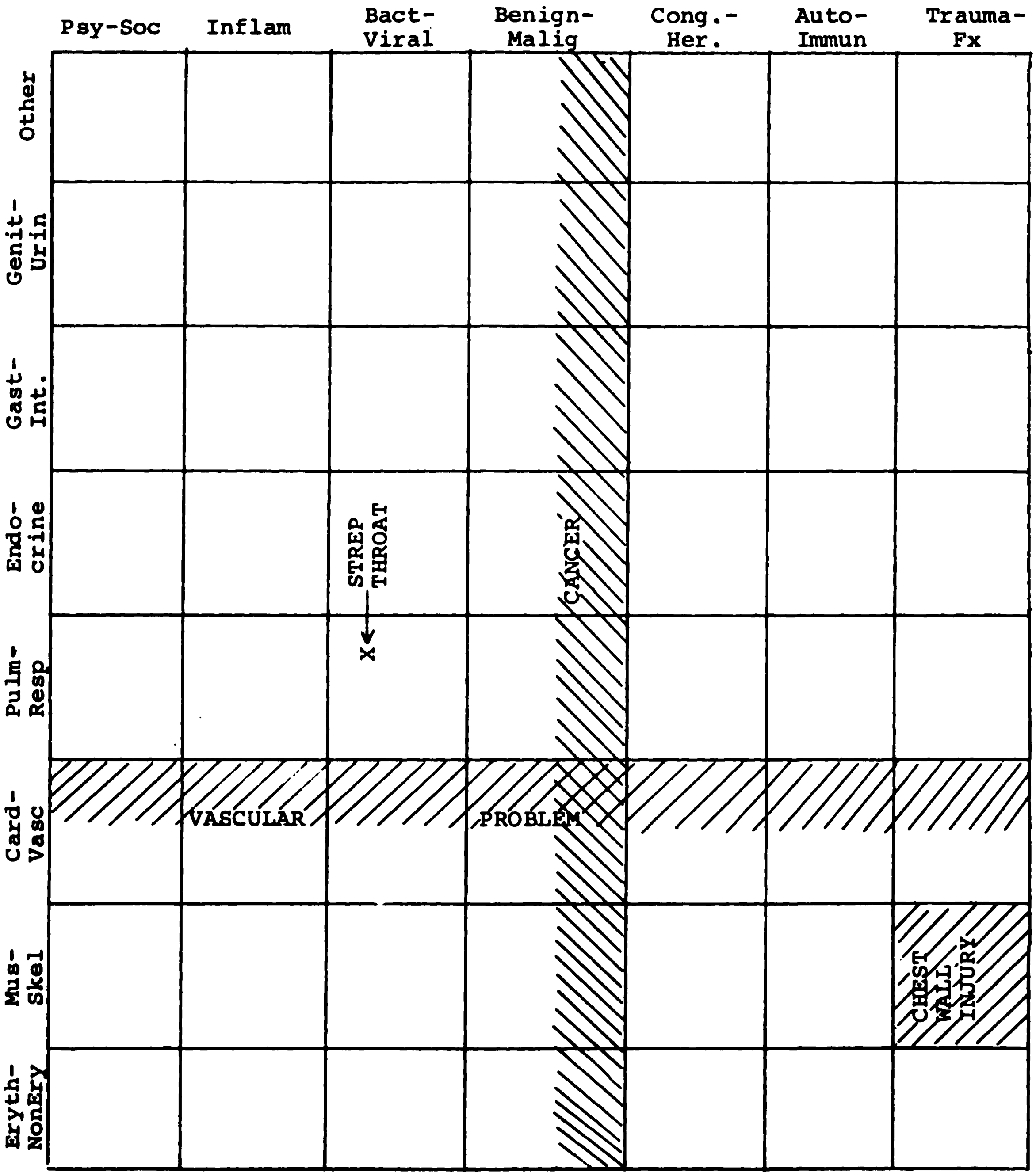

Fig. 2. Disease Process by Organ System Grid[a]

[a] The subject's Early Diagnostic Formulations cover 34 area units of the grid. Specific disease entities such as "strep throat" are scored with X's, each X adding the equivalent of one area unit to the Scope score.

Table 1

Inter-Rater Reliability of Two Independent Raters on the Measure, Scope of Early Diagnostic Formulations for Three Test Problems

Problem	n	Pearson r
Pretest Case 1	32	.91
Posttest Case 3	32	.90
Posttest Case 4	32	.97

Finally, the Scope measure may be considered to have been derived from a test composed of two independent items (cases). A measure of the internal consistency of the test was calculated using a procedure suggested by Hoyt (1967). The internal consistency coefficient obtained was r = .68.

Number of Critically Important Case Findings Elicited

In order to diagnose a case with accuracy, a physician must elicit a significant proportion of diagnostic findings. It is not necessary that all diagnostic findings be elicited since most diseases have more numerous manifestations than are necessary to identify them. In the extreme case, a single pathognomonic finding might be completely unique to a disease and search for additional manifestations for confirmation of the diagnosis would be unnecessary. In previous studies, experienced physicians have elicited approximately 50% of all possible Critical Findings before making a definitive diagnosis (Elstein, 1972). It is presently undetermined how the use of the experimental or idiosyncratic heuristics will effect the number of Critical Findings elicited.

An extensive list of potential findings for the pretest problem and each of the two posttest problems was presented to three physicians familiar with each of the cases. They were instructed to score each of the findings as critically important in arriving at the correct diagnosis (scored ++), somewhat important in arriving at the correct diagnosis (scored +), or noncontributory in arriving at the correct diagnosis (scored 0). In order to assess the stability of the physicians' judgments, correlations between the ratings of the physicians were computed. Using a procedure suggested by Ebel (1967) the inter-rater reliability coefficients of the average ratings of the three physicians on the pretest and two posttest problems were r = .96, r = .94, and r = .93 respectively. The variable of greatest interest in the present study was the number of critically important findings (scored ++) elicited by each subject. Within this category of findings there was 90% agreement among the three judges across the three problems. Disagreements were resolved by consensus of the judges in order to arrive at a final list of findings designated as Critical Findings. The subject's score of the Critical Findings variable was simply the number of such findings elicited during a diagnostic work-up.

Finally, the Critical Findings measure may be considered to have been derived from a test composed of two independent items (cases). A measure of the internal consistency of the test was calculated using a procedure suggested by Hoyt (1967). The internal consistency coefficient obtained was r = .56.

Cost of Information Elicited in the Diagnostic Work-up

It was assumed for the purpose of this study that each piece of information elicited in the course of a search for a diagnosis was associated with some cost. The Cost of the diagnostic work-up was considered to be an additive function of the financial expense incorporating the physician's time, supplies, and equipment; the discomfort and inconvenience to the patient; and the severity and probability of the risk to patient health inherent in the various diagnostic procedures. The financial expense of each procedure was determined by the 1971 Michigan Relative Values Study (Blaine, 1971) of reasonable charges for medical procedures.

In order to determine the relative discomfort and riskiness of diagnostic procedures, five physicians independently rated 25 procedures on a 5-point discomfort scale and on a 5-point scale of concern about risk. (The criterion of concern about risk rather than incidence or prevalence data was used in order to provide a rating that reflected subjective estimates incorporating the aspects of incidence, prevalence, severity, and reversibility of undesirable effects.)

An inter-rater reliability coefficient using a method suggested by Ebel (1967) was computed independently for ratings of discomfort-inconvenience ($r = .88$) and for concern about risk ($r = .56$) among the five physicians. The lower correlation on the risk factor reflects major disagreement on a few procedures; primarily lumbar puncture, bone marrow aspiration, and sigmoidoscopy. Average ratings on the two dimensions were assigned on the basis of the modal physician ratings for each procedure on each of the two scales (see Appendix C for instruments).

In order to combine the independently derived values of financial expense, discomfort, and risk into an overall cost-equivalent for each procedure a variant of a method used by Rubel (1970) was employed. This method essentially entails having physicians assess how many dollars they would be willing to pay to avoid completely the discomfort or risk of the procedure. For example, the question was posed, "Suppose you were the director of a hospital making decisions about drug purchases, and a new radiopaque dye for intravenous pyelograms were available which was guaranteed to be 100% free from adverse reactions. What would you be willing to pay for the new dye, per dose?" The differential between the real price of the currently available dye and the price which the physicians would be willing to pay for the hypothetical dye was assumed to be the dollar equivalent of the physician's concern about the risk inherent in a diagnostic intravenous pyelogram.

Most of the physicians were highly reluctant to commit themselves to exact dollar equivalents; they were much more comfortable providing ranges such as "$30 to $50." The mid-points of the ranges provided by the physicians were used to assign a dollar-equivalent to each point on the 5-point rating scales for discomfort and risk. The dollar-equivalents assigned to each point on the 5-point rating scale were so nearly linear that the equivalents presented in Table 2 were judged to represent the physicians' judgments with no substantial loss of accuracy in this admittedly loose procedure.

Table 2

Rounded Subjective Estimates of Cost-Equivalents for Selected Diagnostic Procedures Rated on Discomfort and Risk

Scale Rating	Discomfort Cost-Equivalent	Risk Cost-Equivalent
1 (minimal)	\$0	\$0
2	\$10	\$20
3 (moderate)	\$20	\$40
4	\$30	\$60
5 (extreme)	\$40	\$80

Substituting the dollar-equivalents for the ratings of each procedure a total Cost value for each procedure could be calculated by the following formula:

$$\text{Cost}_o = \text{financial expense} + \text{dollar-equivalent of rated discomfort} + \text{dollar-equivalent of rated risk}$$

Total Cost of a diagnostic work-up was then calculated by accumulating the Cost of each procedure ordered by a subject in solving each case.

Because the dollar equivalent of discomfort and risk are only rough estimates, the Cost measure may be inherently unstable. In order to assess the stability of the Cost measure, Costs were recomputed using systematically varied coefficients for discomfort and risk. The equations for the alternative Cost estimates were computed as follows:

$Cost_1$ = financial expense + 2x dollar-equivalent of rated discomfort + 2x dollar-equivalent of rated risk

$Cost_2$ = financial expense + .5x dollar-equivalent of rated discomfort + .5x dollar-equivalent of rated risk

$Cost_3$ = financial expense + 2x dollar-equivalent of rated discomfort + .5x dollar-equivalent of rated risk

$Cost_4$ = financial expense + .5x dollar-equivalent of rated discomfort + 2x dollar-equivalent of rated risk

Using the Pearson product moment correlation, each of the alternate total Cost scores for each subject was compared to the original estimate ($Cost_O$). The results of this analysis, reported in Table 3, demonstrate a high degree of stability in the Cost measure over various relative weightings of the three components.

Finally, the Cost measure may be considered to have been derived from a test composed of two independent items (cases). A measure of the internal consistency of the test was calculated using a procedure suggested by Hoyt (1967). The internal consistency coefficient obtained was $r = .47$.

Table 3

Pearson Product Moment Correlations Between Original and Alternate Estimates of Total Cost Scores for Pretest and Posttest Problems

Correlation	n	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)
$r_{cost_o, cost_1}$	32	.984	.994	.999
$r_{cost_o, cost_2}$	32	.993	.994	.999
$r_{cost_o, cost_3}$	32	.997	.996	.999
$r_{cost_o, cost_4}$	32	.992	.995	.998

Accuracy of the Definitive Diagnosis

Following each case subjects were asked to present the details of their diagnosis on a semi-structured short-answer Diagnostic Summary Form (see Figure 3), with instructions to be as specific about their formulation as possible. Formulations of which they were less than certain were to be described as either "possible" or "probable." Indications of whether a formulation was a primary problem, a complication of a primary problem, or secondary problem unrelated to the primary problem were specifically made. This form was the basis for scoring the Accuracy of the diagnosis.

Accuracy scores were calculated using the Diagnostic Accuracy Scoring Form (see Figure 4). Subjects

Case 1

1. **Which of the following** <u>Organ Systems</u> **are, according to the evidence you have collected, implicated in this patient's problem(s)?**

Mark (1) for primary problem, (2) for complication of primary problem, or (3) for unrelated secondary problem.

Organ System	1, 2, or 3	Description of Problem
a. GI Tract		
b. Musculoskeletal	2	myeloma lesions, sixth rib and lumbar spine
c. Cardiovascular	2	hypercholesterolemia
d. Haemopoietic	1, 2	multiple myeloma, anemia
e. Renal		
f. Respiratory		
g. C. N. S.	3	headache - unknown etiology
h. Other		

2. **Which of the following** <u>Disease Processes</u> **are, according to the evidence you have collected, implicated in the patient's problem(s)?**

Mark (1) for Primary problem, (2) for complication of primary problem, or (3) for unrelated secondary problem.

Disease Process	1, 2, or 3	Description of Problem
a. Infection		
b. Neoplasm	1	multiple myeloma
c. Traumatic	2	pathologic fracture of the sixth rib
d. Degenerative Dis.		
e. Congenital		
f. Hereditary Disease		
g. Other		

Fig. 3. Diagnostic Summary Form

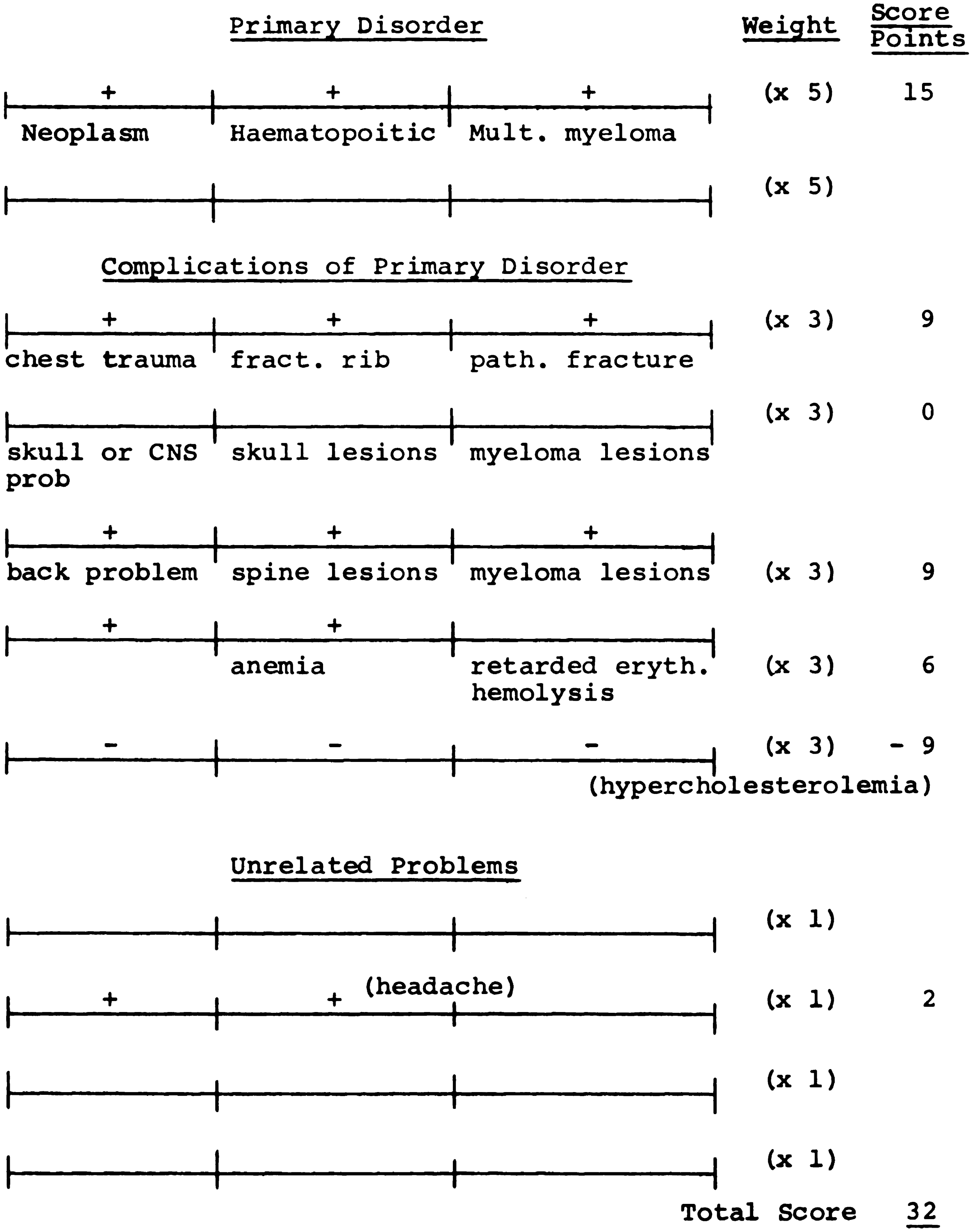

Fig. 4. Diagnostic Accuracy Scoring Form

accumulated Accuracy score points by identifying correctly each disorder. The scoring system rewards greater specificity of diagnosis as well as correct assessment of the relationship of a given disorder to other parts of the diagnosis. Points are subtracted for incorrect inferences.

To determine the inter-rater reliability of the Accuracy scores, three trained judges independently scored each Diagnostic Summary Form for the pretest and two posttest problems. Scores obtained by each judge were then compared using Ebel's (1967) inter-rater reliability procedure. Reliability coefficients for Accuracy scores obtained on the pretest and two posttest problems were $r = .96$, $r = .95$, and $r = .97$ respectively.

It should be noted that each increment in specificity is awarded five points for the primary disorder, three points for each complication, and one point for each unrelated secondary problem. Although it might be generally agreed that these multiples reflect an appropriate priority of importance in the diagnosis, the validity of the particular multiples assigned is questionable. Other equally arbitrary weights might affect the rank ordering of subjects on this measure. In order to assess the stability of the Accuracy measure, scores were computed using three additional sets of relative weights for the categories of primary problem, complications, and unrelated secondary problems. The four sets of weights applied to the categories of diagnosis were as follows:

Accuracy$_1$: primary illness = 5, complications = 3, unrelated secondary problems = 1

Accuracy$_2$: primary illness = 6, complications = 2, unrelated secondary problems = 1

Accuracy$_3$: primary illness = 8, complications = 3, unrelated secondary problems = 1

Accuracy^4: primary illness = 4, complications = 3, unrelated secondary problems = 1

Pearson product moment correlations were computed between scores obtained using the original set of weights (Accuracy$_1$) and each of the alternative sets of weights. The results of this analysis, reported in Table 4, indicate that Accuracy scores are highly stable over changes in weights.

Table 4

Correlations Among Scores of Diagnostic Accuracy Obtained by Four Systems of Weights

Correlation	n	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)
r_{acc_1, acc_2}	32	.996	.983	.986
r_{acc_1, acc_3}	32	.949	.990	.979
r_{acc_1, acc_4}	32	.999	.998	.993

Finally, the Accuracy measure may be considered to have been derived from a test composed of two independent items (cases). A measure of the internal consistency of the test was calculated using a procedure suggested by Hoyt (1967). The internal consistency coefficient obtained was r = .25.

Summary of Reliability Studies on Dependent Variables

The reliability of the dependent measures was investigated in several ways. First, studies of agreement among experts were required in order to develop scoring keys on the Critical Findings and Cost variables. Second, studies of the stability of subject scores over transformations in scoring rules were appropriate for the variables of Cost and Accuracy. Third, studies of inter-rater reliability on the scoring of subjects' performance were required for the Scope and Accuracy measures. Fourth, studies of consistency of subject's performance across problems were appropriate for all four variables. Considering each measure in turn, the results of the studies are summarized in the following paragraphs.

The Scope scoring format was developed rationally and required no empirical judgments for the development of scoring keys. Substantial judgment by scorers was required to score a subject's performance and inter-rater reliability was high (r = .90). Internal consistency on the Scope measure was r = .68 (Hoyt).

The Critical Findings variable required a key developed by a panel of three physicians. These physicians achieved reliabilities above .90 in their judgments of degree of importance and an average of approximately 90% agreement on the Critical Findings in the posttest cases. Once the key was developed for the Critical Findings variable, scoring was completely objective and no investigation of inter-rater reliability was considered appropriate. Internal consistency for the Critical Findings measure was r = .56 (Hoyt).

The development of the Cost measure required consistency of expert judgment for patient discomfort (r = .88) and risk (r = .56). Applying various relative weights to the components of expense, discomfort, and risk, differences in the aggregate Cost scores were found to be negligible. Objective scoring procedures on the Cost variable eliminated the need for inter-rater reliability studies. The internal consistency of the Cost measure was r = .47 (Hoyt).

The Accuracy measure did not require empirical keying, but the relative weights applied to subscores were arbitrary. Stability of the Accuracy scores over transformations of subscore weights was extremely high. The mean inter-rater reliability in the scoring of subject's performance was also quite high (r = .91). Consistency of subjects performance on the Accuracy measure across problems was disappointing (r = .25, Hoyt).

Statement of the Hypotheses

The research questions posed in Chapter I of this study may now be stated as formal null hypotheses to be tested by the statistical procedures reported in Chapter IV. The hypotheses are stated in the null form rather than as research hypotheses for the following reasons. First, for several of the hypotheses no differences were expected. Second, the empirical literature and theory with respect to each of the hypotheses was either inadequate or contained discrepancies which made the basis for prediction quite speculative. Third, the writer, in his role as experimenter was in a position to influence the performance scores of subjects. In order to maintain experimental neutrality in his interpretations of information-requests and the dispensing of data, the experimenter needed to bear in mind the variety of equally interesting possible relationships between the various treatment groups and the various dependent measures.

Hypothesis 1:

There is no mean difference among the scores of fourth-year medical students receiving different combinations of heuristic training and prompting on any of the following variables defined in this study: Scope of Early Diagnostic Formulations, Number of Critical Findings, Cost of the Diagnostic Work-up, Accuracy of the Definitive Diagnosis.

Hypothesis 2:

There is no mean difference among the scores of fourth-year medical students enrolled at the University of Michigan and scores of medical students enrolled at Michigan State University on any of the following variables defined in this study: Scope of Early Diagnostic Formulations, Number of Critical Findings, Cost of the Diagnostic Work-up, Accuracy of the Definitive Diagnosis.

Hypothesis 3:

There are no interactions between the combination of heuristic training and prompting and the medical school of enrollment on any of the following variables defined in this study for fourth-year medical students: Scope of Early Diagnostic Formulations, Number of Critical Findings, Cost of the Diagnostic Work-up, Accuracy of the Definitive Diagnosis.

Hypothesis 4:

There are no significant correlations among scores obtained by fourth-year medical students on any of the following variables defined in this study: Scope of Early Diagnostic Formulations, Number of Critical Findings, Cost of the Diagnostic Work-up, Accuracy of the Definitive Diagnosis.

Experimental Design

The hypotheses of this study were tested by means of four two-way analyses of covariance. The pretest score on each of the four dependent measures was used as the covariate for its respective analysis. The factorial design of the study is displayed graphically in Figure 5.

Multivariate analysis of covariance was considered as an alternative to the univariate analyses, but

	Experimental Treatments			
	1	2	3	4
Medical School 1	n = 4	n = 4	n = 4	n = 4
Medical School 2	n = 4	n = 4	n = 4	n = 4

Fig. 5. Factorial Design for Analysis of Treatment Group and Medical School Effects on Each of Four Dependent Variables

it was judged that simultaneous probability statements about the four dependent variables could not be meaningfully interpreted.

Because the present study can be classified as an early attempt to investigate the effects of training on a highly complex task, clear-cut results were a highly optimistic expectation. In such high-risk experimentation trends in data, new hypotheses, and better understanding of the variables are more usually found than clear and important differences between groups. For this reason, the experimental design of the study includes a two-stage decision rule based upon the statistical significance levels of the various experimental contrasts.

Stage 1. Because multiple univariate analyses have been calculated, the alpha level for the entire experiment was inflated. Therefore decisions to reject the null hypotheses should be based on a conservative decision rule. In this study the null hypotheses will be rejected at or below the .01 level of significance.

Stage 2. The dismissal of promising trends which did not reach the critical alpha level is judged to be an inefficient use of data, particularly in exploratory research. Consequently, a second decision rule--the decision to retain the alternate hypothesis as a highly promising hypothesis worthy of further and more refined investigation--will be stated. The alpha level selected for the second decision rule was the .10 level of significance.

The research questions involving relationships between dependent measures rather than contrasts between experimental groups were computed using Pearson product moment correlations.

CHAPTER IV

RESULTS

The results of the experiment are reported in this chapter. For the factorial analysis of each of the four dependent measures, the mean and standard deviations on each measure and its respective analysis of covariance table are presented. In addition to the factorial analysis of the dependent measures, suggestive patterns among dependent measures and correlational results are reported. Several incidental analyses conclude this chapter.

Factorial Analysis of Dependent Measures

Scope of the Early Diagnostic Formulations (Scope)

The maximum possible Scope score was 224. Post-test scores ranged from 12 to 98 with a grand mean of 54 and a standard deviation of 31.1. Treatment group means and standard deviations on the Scope measure are reported in Table 5. Using the Scope score on the pre-training problem (case 1) as a covariate and the mean Scope score

Table 5

Treatment Group Means and Standard Deviations on Scope Measure

Treatment Group	n	Mean	Adjusted Mean	Standard Deviation
1	8	68.25	68.54	27.05
2	8	52.50	52.99	25.18
3	8	60.00	59.74	30.37
4	8	47.50	46.98	19.10

on the two posttest problems (cases 3 and 4) as the dependent variable, a two-way analysis of covariance for differences among the four treatment groups and for difference between medical schools revealed no significant difference between the treatment groups or medical schools on this variable. Results are reported in Table 6.

Since no significant differences were found, post hoc procedures were considered inappropriate.

Number of Critical Findings Elicited (Critical Findings)

The maximum possible Critical Findings score was 19 points. Scores ranged from 6 to 18 points with a grand mean of 10.5 and a standard deviation of 3.18. Treatment group means and standard deviation on the Critical Findings measure are reported in Table 7.

Table 6

Two-Way Analysis of Covariance on Scope of Early Diagnostic Formulation Score

Source of Variation	SS	df	MS	F	P
T: Treatments	1950	3	650	.90	.46
S: Schools	1271	1	1271	1.75	.20
T x S Interaction	1197	3	399	.55	.65
Error	16606	23	722		
Total	21024	30			

Table 7

Treatment Group Means and Standard Deviations on Critical Findings Measure

Treatment Group	n	Mean	Adjusted Mean	Standard Deviation
1	8	10.75	11.15	2.50
2	8	10.25	11.02	1.30
3	8	11.87	11.13	3.10
4	8	10.75	10.32	3.37

Using the Critical Findings scores on the pretest problem as a covariate and the mean Critical Findings score on the two posttest problems as the dependent variable, a two-way analysis of covariance for differences between treatment groups and differences between medical schools revealed no significant differences on the number of Critical Findings elicited. Results of the analysis are reported in Table 8.

Table 8

Two-Way Analysis of Covariance on Number of Critical Findings Elicited

Source of Variation	SS	df	MS	F	P
T: Treatments	3.36	3	1.12	.14	.93
S: Schools	5.99	1	5.99	.77	.39
T x S Interaction	3.66	3	1.22	.16	.92
Error	184.00	23	8.00		
Total	197.01	30			

Since no significant differences were found post hoc procedures were considered inappropriate.

Cost of Diagnostic Work-up (Cost)

The maximum possible Cost score is indeterminate since subjects were free to request as much information as they desired and were permitted to repeat tests when they questioned the reliability of the information received. Cost scores ranged from $63 to $1179 with a grand mean of $287 and a standard deviation of $215.90. Treatment group means and standard deviations on the Cost measure are reported in Table 9.

Table 9

Treatment Group Means and Standard Deviations on Cost Measure

Treatment Group	n	Mean	Adjusted Mean	Standard Deviation
1	8	185.87	205.35	75.94
2	8	260.37	251.24	122.61
3	8	370.75	371.13	342.19
4	8	330.75	320.03	165.13

Using the Cost score on the pretest problem as a covariate and the mean Cost score of the two posttest problems as the dependent variable, a two-way analysis of covariance for differences among treatment groups and difference between medical schools revealed no significant differences on the Costs of the Diagnostic Work-ups. Results of this analysis are reported in Table 10.

Since no significant differences were found post hoc comparisons were considered inappropriate.

Table 10

Two-Way Analysis of Covariance on Cost of Diagnostic Work-up

Source of Variation	SS	df	MS	F	P
T: Treatments	124569	3	41523	.86	.48
S: Schools	7519	1	7519	.16	.70
T x S Interaction	130419	3	43473	.90	.46
Error	1110486	23	48282		
Total	1372993	30			

Accuracy of the Definitive Diagnosis (Accuracy)

The maximum possible Accuracy score was 55 points. Scores ranged from 24 to 50 with a grand mean of 40.6 and a standard deviation of 6.7. Treatment group means and standard deviations on the Accuracy measure are reported in Table 11.

Using the subjects Accuracy score on the pretest problem as a covariate and the mean of the Accuracy scores on the two posttest problems as a dependent variable, a two-way analysis of covariance for differences among treatment groups and differences between medical schools was computed. Statistical significance was approached for the treatment effect ($p < .07$) but did not reach the conservative critical value adopted ($P < .01$) to compensate for an inflated alpha level in the overall analysis. Results of the covariance analysis are reported in Table 12.

Table 11

Treatment Group Means and Standard Deviations on Accuracy Measure

Treatment Group	n	Mean	Adjusted Mean	Standard Deviation
1	8	45.87	45.83	2.89
2	8	41.00	41.03	8.28
3	8	36.75	36.82	6.29
4	8	40.00	39.94	6.58

Table 12

Two-Way Analysis of Covariance on Accuracy of the Definitive Diagnosis

Source of Variation	SS	df	MS	F	P
T: Treatments	434.34	3	144.78	2.69	.07
S: Schools	15.73	1	15.73	.37	.55
T x S Interaction	63.54	3	21.18	.50	.69
Error	1237.86	23	53.82		
Total	1787.47	30			

Following the two-stage decision rule adopted for interpretation of results in Chapter III, the possibility of treatment effects on scores of diagnostic Accuracy is retained as a hypothesis worthy of further investigation.

Since the omnibus F test on the Accuracy measure failed to yield statistically significant differences, post-hoc analyses of differences between specific treatments were inappropriate. The trends suggest, however, that unprompted subjects (groups 2 and 4) have similar Accuracy scores whether using their idiosyncratic heuristics or previously trained to use and encouraged to use the experimental heuristics. Of greater interest is the possibility that systematic use of the experimental set of heuristics by group 1 subjects may have improved their diagnostic Accuracy scores, while systematic use of the idiosyncratic heuristics by group 3 subjects may have adversely influenced their diagnostic Accuracy scores.

Since performance by subjects varied widely across cases, it was felt that deriving dependent measures from mean performance scores on the two posttest cases might have obscured some relationships peculiar to individual cases. Therefore analyses of covariance were recomputed on the four dependent variables with scores derived independently from cases 3 and 4 rather than

from their mean. In none of the eight resulting analyses did mean differences approach statistical significance.

Correlational Analyses Among Dependent Variables

Each of the foregoing analyses has dealt with differences between the experimental groups on the dependent variables. Correlations among scores of all subjects were also calculated in order to detect relationships among various performance measures and to assess the stability of subjects' performance across the problem cases. In this section the results of several correlational analyses are reported in order to answer these further research questions. Some new variables incidental to the main analysis are introduced and analyzed in this section. Significance levels are reported for each of the numerous correlations in the following sections. Readers should be cautioned that the probability is quite high that several of the correlations reached statistical significance purely by chance.

Consistency of Performance Across Problems

In addition to the dependent measures previously analyzed, other measures of problem-solving performance were available. The additional measures included clinical findings elicited which were of less than critical importance (defined as Moderately Important Findings and Noncontributory Findings) and the separate components of the aggregate Cost measure (including financial expense and the dollar equivalents of diagnostic discomfort and risk in the medical history, physical examination, and laboratory procedure section of the work-up).

Pearson product moment correlations were computed for each of the additional measures between problem pairs. These correlations, presented in Table 13, reflect the consistency of subjects' behavior on problems of similar format but different content.

Table 13 reveals an inconsistent pattern for most of the performance measures across problems. Measures 2 through 6, however, are significantly correlated across problems and all deal with what might be called the extensiveness of the search for data. Subjects exhibited characteristic styles of data collecting in the history and physical examination portions of their work-ups. Some preferred to be systematic and thorough, others preferred to obtain only a brief history and a selective physical examination. This difference in style is important to the questions of this study because it reflects the extent to which various medical students are operating in the stepwise approach emphasized in their training or in a hypothesis-guided approach. The best indicator of this approach differential is probably a composite of the measures Financial Expense of History (a direct

Table 13

Pearson Product Moment Correlations on Performance Measures Across Problems for All Subjects (N = 32)

Performance Measures	Correlates		
	$r_{3,4}$	$r_{1,3}$	$r_{1,4}$
1. Scope of Early Formulations	$.45^b$	.06	.05
2. Critical Findings Elicited	$.42^b$	.23	$.39^a$
3. Mod. Import. Findings Elicited	$.57^c$	$.45^b$	$.58^c$
4. Noncontrib. Findings Elicited	$.83^c$	$.50^b$	$.60^c$
5. Financial Expense of History	$.76^c$	$.45^b$	$.60^c$
6. Financial Expense of Physical Exam	$.61^c$	$.32^a$	$.34^a$
7. Financial Expense of Lab Tests	$.37^b$	-.03	.17
8. Discomfort of Physical Exam	.23	-.16	-.21
9. Discomfort of Lab Tests	.25	.03	.22
10. Risk of Physical Exam	1.00^c	-.16	1.00^c
11. Risk of Lab Tests	.15	.16	.21
12. Cost of Work-up	$.35^a$	.02	.25
13. Accuracy of Diagnosis	-.06	.03	-.19

$^a p < .05$

$^b p < .01$

$^c p < .001$

function of the number of historical inquiries made by the subject) and Financial Expense of the Physical Examination (a direct function of the estimated time to perform each segment of a physical exam). This composite will be labeled a "Thoroughness" measure and will be included in subsequent correlational analyses.

Two other performance measures in Table 13 deserve mention. First, the peculiar pattern of correlations on measure 10, Risk of Physical Exam, arose because only one procedure of physical examination, the sigmoidoscopy, was associated with a risk factor. The correlations on this measure can therefore be dismissed as an artifact. Second, the nonsignificant correlations on measure 13, Accuracy of Diagnosis, indicates that this important outcome may be more a function of the students' content specific knowledge of medicine than his approach to problem solving. More will be said about this point in a later section of this chapter.

Relationship Between Diagnostic Cost and Accuracy

Figure 6 depicts the adjusted mean scores for each of the four heuristic treatment groups on the diagnostic Cost and Accuracy measures.

Figure 6 suggests an inverse relationship between diagnostic Cost and Accuracy but is misleading not only because of failures of the measures to reach statistical

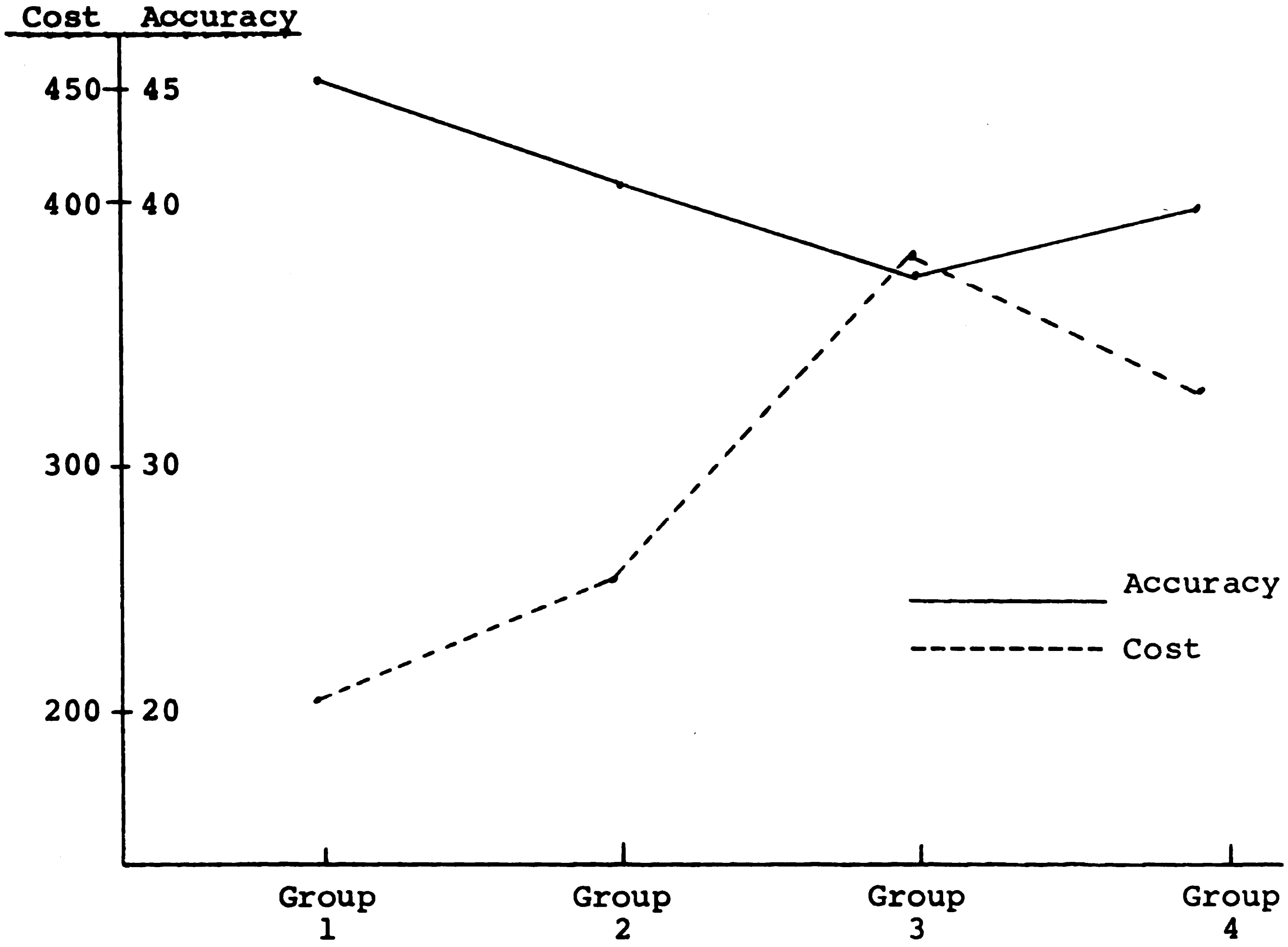

Fig. 6. Adjusted Mean Scores of Treatment Groups on Cost and Accuracy

significance, but also because inferences made would refer only to groups and not to individual subjects.

Correlations between diagnostic Cost and Accuracy on each of the pretest and posttest problems using subjects as the unit of analysis, support the conclusion that there is no significant linear relationship between these two measures of diagnostic effectiveness. Results of the correlational analysis are reported in Table 14.

Table 14

Pearson Product Moment Correlations Between Cost of Diagnostic Work-up and Accuracy of Diagnosis on Individual Problems for All Subjects (N = 32)

Correlates	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)
$r_{\text{cost, accuracy}}$	.13	.17	-.05

$^a p < .05$

The Cost of the diagnostic work-up and the Accuracy of the definitive diagnosis are considered to be outcome measures of diagnostic problem solving and should be considered separately from all of the remaining measures, which are intended to assess processes utilized in achieving the diagnostic outcomes. The relationship between the two outcomes of Cost and Accuracy may be complex. One might expect more extensive investigation of a problem by subjects and concomitantly higher costs to yield more accurate diagnosis. Conversely, higher Cost might occur when a subject requests several noncontributory but costly laboratory tests in pursuit of an inaccurate diagnosis. Finally, there may be no significant linear relationship between Cost and Accuracy because a costly and inefficient search may eventually lead to as accurate a diagnosis as a more efficient search.

The form of nonlinear relationship between diagnostic Cost and Accuracy was investigated in a preliminary way by examination of scatter plots of these measures on each of the three problems. The scatter plots, included in Appendix D, indicate a ceiling effect on the Accuracy variable in case 1 and no discernable relationship among the Cost and Accuracy measures on either of the two posttest problems.

Relationship Between Diagnostic Cost and Selected Process Measures of Problem Solving

The Cost of a diagnostic work-up may be a function of any of several measures of behavior exhibited by subjects as they proceed through medical problems. Table 15 presents the correlations between diagnostic Cost outcomes and the process variables of Scope of early diagnostic formulations, Thoroughness of history and physical examination, number of Critical, Moderately Important, and Noncontributory findings elicited.

Table 15

Pearson Product Moment Correlations Between Cost of Diagnostic Work-up and Selected Process Measures on Individual Problems for All Subjects (N = 32)

Correlates	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)
$r_{\text{cost, scope}}$	.20	.29	.17
$r_{\text{cost, thoroughness}}$	$.41^b$	$.32^a$	$.53^c$
$r_{\text{cost, critical find.}}$	$.43^b$	$.39^b$	$.61^c$
$r_{\text{cost, mod. import. find.}}$	$.51^b$	.18	$.52^b$
$r_{\text{cost, noncontrib. find.}}$	$.53^c$	$.34^a$	$.71^c$

$^a p < .05$

$^b p < .01$

$^c p < .001$

The pattern of correlations in Table 15 indicates that higher Cost is associated with greater Thoroughness and with more numerous findings of critical importance, moderate importance, and noncontributory importance. No significant relationship was detected between Cost and the Scope of the early problem formulation. The significant correlations found were entirely expected since the

Thoroughness measure was actually a component of the Cost measure and each finding was directly associated with some defined unit Cost incurred in its elicitation.

Relationship Between Diagnostic Accuracy and Selected Process Measures of Problem Solving

As with the outcome of Cost, Accuracy of the definitive diagnosis may be a function of any of several measures of behavior exhibited by subjects as they proceed to solutions of diagnostic problems. Table 16 presents the correlations between the Accuracy of the definitive diagnosis and the process variables of Scope of the Early Diagnostic Formulation, Thoroughness of history and physical examination, and number of Critical, Moderately Important, and Noncontributory Findings elicited.

The results reported in Table 16 suggest that diagnostic Accuracy scores are positively correlated with the number of Critical Findings elicited but not with the Scope of Early Diagnostic Formulations as might be expected from Barrow's and Bennett's observations (1972), nor with the Thoroughness of the data search as might have been anticipated by those endorsing the stepwise approach to diagnosis. The correlation between Critical Findings elicited and diagnostic Accuracy seems reasonable; perhaps most surprising is the finding that on the complex problems presented in this study performance on the

Table 16

Pearson Product Moment Correlations Between Accuracy of the Definitive Diagnosis and Selected Process Measures on Individual Problems for All Subjects (N = 32)

Correlates	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)
$r_{\text{accuracy, scope}}$	-.02	.29	-.05
$r_{\text{accuracy, thoroughness}}$	.49[b]	.08	-.14
$r_{\text{accuracy, critical find.}}$	.34[a]	.66[c]	.01
$r_{\text{accuracy, mod. import. find.}}$	.49[b]	.07	-.12
$r_{\text{accuracy, noncontrib. find.}}$	.30[a]	.05	-.22

[a] p .05

[b] p .01

[c] p .001

Critical Findings measure accounted for only about 13% of the variance on the Accuracy measure.

Relationships Between Process Measures

The process measures of Scope of the Early Diagnostic Formulations, Thoroughness of history and physical examination, and Critical, Moderately Important, and Noncontributory Findings elicited have been examined with respect to their correlations with each other. Since performance on these measures is relatively stable across problems (see Table 13) the mean correlations between

measures across problems is presented in Table 17. The mean correlations were computed using the r to Z transformation.

Table 17

Mean Pearson Product Moment Correlations (After r to Z Transformation) Between Selected Process Measures Across Problems for All Subjects (N = 32)

Variables	1	2	3	4	5
1. Scope	1.00				
2. Thoroughness	.29	1.00			
3. High Import. Find.	.29	.62[c]	1.00		
4. Mod. Import. Find.	.31[a]	.80[c]	.58[c]	1.00	
5. Noncontrib. Find.	.26	.89[c]	.58[c]	.74[c]	1.00

[a] p .05

[b] p .01

[c] p .001

The results reported in Table 17 demonstrate a strong but not surprising positive relationship between Thoroughness of data search and the number of Critical, Moderately Important, and Noncontributing Findings elicited. The lack of a significant correlation between the Scope score and the Thoroughness score suggests that early general hypotheses are not associated with a more thorough search than early specific hypotheses.

Incidental Analyses

Relationships Between Medical College Admissions Test Scores and Scores on Cost of the Diagnostic Work-up and Accuracy of the Definitive Diagnosis

The Medical College Admissions Test (MCAT) is a standardized achievement test widely used in the selection of applicants to medical school. The test is composed of four parts yielding scores on verbal knowledge, quantitative knowledge, science knowledge, and knowledge of general information. The MCAT has recently come under fire from critics who argue that the test has virtually no validity beyond its minimal ability to predict grade-point average in the first year of medical school and should be replaced by an instrument which would predict clinical competence; the ability to solve diagnostic problems and to make management decisions. The MCAT has not, however, been given a fair trial as a predictor of clinical competence since the latter ability is typically evaluated only by the subjective and global impressions of clinical faculty. Such impressions are perhaps unduly influenced by personality variables which may shape the relationship between student and faculty and enter into subjective clinical evaluations.

While the performance measures of this study are intended to tap only a few aspects of clinical competence, it seems appropriate to include the correlations between MCAT scores and outcome measures of diagnostic problem-solving performance. Table 18 presents correlations between MCAT scores and scores of Cost of the diagnostic work-up. Table 19 presents correlations between MCAT scores and scores of Accuracy of the definitive diagnosis. The MCAT tests used in this analysis were administered to the subjects prior to their entry into medical school approximately three years earlier.

Table 18

Pearson Product Moment Correlations Between Scores on the Medical College Admissions Test and Cost of the Diagnostic Work-up on Three Problems for All Subjects (N = 32)

MCAT Scores	Cost of Diagnostic Work-Up			
	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)	Mean (Cases 1, 3, 4)
Verbal	-.11	-.09	-.28	-.16
Quantitative	.08	.03	-.05	.02
General Information	.03	.11	.03	.06
Science	-.07	-.18	-.25	.17
Total	-.02	-.05	-.21	-.09

$^{a}p < .05$

The correlations reported in Tables 18 and 19 appear to be randomly distributed about the zero point. The single correlation to reach statistical significance at the .05 level (r = -.30) is readily explained on the basis of chance alone. Thus, no relationship has been found between MCAT scores and Cost or Accuracy of diagnostic problem solving.

Table 19

Pearson Product Moment Correlations Between Scores on the Medical College Admissions Test and Accuracy of the Definitive Diagnosis on Three Problems for All Subjects (N = 32)

MCAT Scores	Accuracy of the Definitive Diagnosis			
	Pretest (Case 1)	Posttest (Case 3)	Posttest (Case 4)	Mean (Cases 1, 3, 4)
Verbal	-.20	.26	.20	.09
Quantitative	-.07	.12	-.17	-.12
General Information	-.12	.08	.01	-.01
Science	-.30[a]	.15	-.20	-.12
Total	-.25	.23	-.06	-.03

[a] $p < .05$

Evaluation of the Experimental Heuristics by Subjects

The subjects of this study were beginning their fourth year of medical school and had been previously engaged in extensive supervised clinical work. It was therefore possible that they were already quite aware of the experimental heuristics and might have learned to apply them through clinical experience. Conversely, some of the heuristics might be in opposition to their previous training and experience, making their application to the problems of the study seem difficult and unnatural. In order to assess the subjects' familiarity with the experimental heuristics and the consistency of the heuristics with their clinical training, the sixteen subjects who were given training on the experimental heuristics were asked to complete a multiple choice questionnaire assessing the heuristics (see Appendix C). The familiarity dimension of the questionnaire referred to the subject's exposure to heuristics substantially equivalent to the experimental heuristics. The consistency dimension referred to the subject's opinion of the extent to which faculty members would endorse each heuristic. The frequency of response to each questionnaire item is summarized in Table 20.

Content Analysis of Idiosyncratic Heuristics

The eight subjects of experimental group 3 were each asked to generate a list of four to six rules of thumb which they had learned in their clinical training and had found helpful in solving diagnostic problems. These subjects generated a total of 44 heuristics. The idiosyncratic heuristics can be divided into two main groups: (1) those referring to data collection and (2) those referring to data interpretation. Five of the heuristics generated were considered to be uninterpretable within this framework. Data collection

Table 20

Frequency of Subject Responses on Perceived Familiarity and Consistency with Clinical Training of Experimental Heuristics (N = 16)

	Familiar Consistent	Not Familiar Consistent	Not Familiar Indifferent	Controversial	Not Familiar Inconsistent
Heuristic 1					
Sch. A	2	5	0	0	1
Sch. B	6	1	0	0	1
Heuristic 2					
Sch. A	1	2	3	1	1
Sch. B	4	3	0	1	0
Heuristic 3					
Sch. A	1	5	0	1	1
Sch. B	5	2	0	0	1
Heuristic 4					
Sch. A	0	7	0	0	1
Sch. B	4	4	0	0	0
Heuristic 5					
Sch. A	4	2	0	2	0
Sch. B	3	3	1	1	0
All Heuristics					
Sch. A	8	21	3	4	4
Sch. B	22	13	1	2	2
Combined Schools	30	34	4	6	6

heuristics varied from general approaches to reminders about making specific patient inquiries. The data interpretation heuristics were all rather general with two exceptions referring to the differential diagnosis of anemia and exhaustion. The data interpretation heuristics could not be reliably differentiated as applying to either hypothesis generation or to hypothesis testing. Some referred specifically to both aspects, others such as "keep an open mind" might refer to either the generation or the testing of hypotheses. The idiosyncratic heuristics are presented in Table 21. Where several heuristics conveyed essentially the same message they were grouped together and frequencies were reported. The original idiosyncratic heuristics of each subject are reported verbatim in Appendix C.

Summary of Results

None of the formal null hypotheses of treatment group differences could be rejected at the conservative .01 level of significance adopted for the factorial analysis. Results of the factorial analysis of the Diagnostic Accuracy measure, however, indicated a significant treatment effect at the .07 level; suggesting the possibility that Accuracy scores were differentially influenced by particular combinations of heuristic training and prompting. An important unanticipated

Table 21

Content of Idiosyncratic Heuristics Generated by Eight Subjects

		Frequency
Data-Gathering Heuristics--General		
1.	Gather a complete history and physical including confirmation from other sources when patient is unreliable.	6
2.	Listen carefully to what the patient says.	4
3.	Collect data system by system.	2
4.	Include pertinent review of systems of problems suspected within review of present illness.	1
5.	Obtain history in chronological form.	1
6.	Order lab tests in a noninterfering sequence.	1
7.	When confused, go to review of systems.	1
Data Gathering--Specific		
Include the following inquiries routinely:		
8.	Family history.	1
9.	Examination of hands and skin.	1
10.	Rectal digital examination.	1
11.	Urinalysis (3), complete blood count (2), VDRL (1), TB test (1), chest film (1).	3
Data Interpretation--General		
12.	Think of common things first.	2
13.	Use common sense.	1
14.	Keep an open mind.	1
15.	Think by systems.	1
16.	Think by differential diagnosis.	1
17.	Apply the principle of parsimony.	1
18.	Go from general differential diagnosis in history to specifics in physical and lab.	1
19.	Go from general to specific tests in laboratory diagnosis.	1
20.	Maximize thought processes before laboratory work.	1

Table 21 (Continued)

	Frequency
Data Interpretation--General (Continued)	
21. Pay attention to detail but don't get lost in detail.	1
22. Evaluate data as primary or secondary.	1
23. Clinical findings should be regarded as more reliable than lab findings.	1
24. When diagnosis seems apparent, don't completely rule out other etiologies.	1
25. Evaluate patient's response to therapy for diagnostic clues.	1
Data Interpretation--Specific	
26. Differentiate anemias as (1) blood loss, (2) hemolytic, or (3) production deficiency.	1
27. In exhaustion, remember muscle strength and food fads.	1
Uninterpretable Heuristics	
28. The sin of commission is worse than the sin of omission or vice versa.	1
29. Think first--don't spout the first thing that comes into your head.	1
30. Physical exam.	1
31. Pertinent lab results.	1
32. Treatment.	1
	44

finding was the enormous variability among scores on the Cost variable. This finding perhaps more than any other reported in this study should interest those charged with the clinical education of physicians.

Correlational analyses demonstrated no linear relationships between the two measures of diagnostic outcome--Cost and Accuracy. Both of the outcome measures were correlated with several measures of the diagnostic process. Among the process measures the Number of Critical, Moderately Important, and Noncontributory Findings were each correlated with Cost of the work-up. Only the Critical Findings variable, however, was related to Diagnostic Accuracy. The Scope of Early Diagnostic Formulations appeared to bear no significant relationship to either Cost or Accuracy.

Correlations computed to assess the stability of subjects' behavior across problems yielded generally significant though low to moderate correlations for the process measures. A conspicuous lack of internal consistency was noted for the Accuracy variable.

Correlations among the several process measures indicated that the Thoroughness of the medical history and physical examination had little or no relationship to the Scope of the Early Hypotheses. Correlations between the Medical College Admission Test and the outcome measures provided no indication that the MCAT had predictive validity in the domain of diagnostic problem solving.

CHAPTER V

DISCUSSION

In his role as the experimenter, the writer spent approximately 130 hours in individual problem-solving sessions with the 32 experimental subjects. In addition considerable time was spent with medical students and physicians in the pilot study phase of the project. Observations and anecdotes were recorded during these encounters which may aid in the interpretation of the previously reported results. The collection of these notes was fortuitous rather than systematic, however, so the interpretations presented should be viewed as speculative and tentative but plausible explanations of the results obtained.

The single most striking impression of the experimenter on completion of the data collection phase of the study was the extreme variability in virtually every dimension of problem-solving behavior investigated. The theme of variability was manifested in great heterogeneity among subjects on all variables, large variability in performance by the same subject on different problems and a noticeable lack of standardization in the approach to the same problem by different subjects. The great variability in both processes and outcomes makes the use of terms such as "trends" and "tendencies" very dangerous because generalizations beyond narrow limits of experience are not warranted. Hence, a good deal of the interpretation of results must take place at a microscopic rather than macroscopic level.

Discussion of Schools Variable

As previously mentioned the sample for the study was drawn from two medical schools. Students were similar at medical school entry in terms of their pre-medical academic achievement as measured by the MCAT but dissimilar in some aspects of subsequent training. The factor of Medical School was included in the design of the study to increase the external validity of the study. As anticipated no significant differences between medical schools on any of the variables was found. It can be concluded that the results of the experimental treatment variable are generalizable to at least two Michigan medical schools having students of similar previous achievement but dissimilar curricula. In further interpretation only the comparison among experimental treatment groups will be investigated and these comparisons should be understood to apply to both medical schools in the study.

Discussion of Scope Results

Subjects in treatment groups 1 and 2 had been trained to employ the hypothesis specificity heuristic among others. This heuristic was, in fact, a direct instruction to review the Scope of each hypothesis generated and to alter the hypotheses to make them more closely correspond to the currently available data base. Thus, the difference between treatment group means on the Scope measure can be said to be a direct reflection of the extent to which subjects understood and applied this particular heuristic in the posttest cases. In the training phases all subjects were quickly able to grasp the concept and rationale for keeping the Scope of their diagnostic formulations consistent with the available supporting data. Subjects were able to generate examples of inappropriately narrow or broad hypotheses from their own experience. Despite their conceptual understanding, subjects varied greatly in their ability to apply this heuristic in the problem-solving posttest. This failure is seen as the primary reason for lack of significant differences among treatment groups on the Scope score.

The different kinds of responses to the hypothesis specificity heuristic are illustrative of problems that may occur generally in the application of heuristic suggestions to problem solving. First, there were some subjects who were either unwilling or unable to alter the statement of their hypotheses after review of the hypothesis specificity heuristic. Under conditions where an unjustifiably specific hypothesis had been generated, these subjects indicated no recognition of the discrepancy between the specificity of the hypothesis and the deficiency of the supporting data base. A portion of such subjects appeared to be too involved to achieve a perspective of the problem. As one student remarked afterwards, he "couldn't see the forest for the trees." It would seem that awareness of situations in which particular heuristics are applicable is a skill that does not appear automatically and should be adequately accounted for in heuristic training.

Another portion of subjects unable to alter their hypotheses simply appeared reluctant to expend the cognitive energy on refinements of hypotheses while they felt they were making good progress toward the solution. Voice intonation, impatient glances, and other nonverbal cues conveyed the message that the alteration of hypotheses was considered to be an unwelcome distraction from some subjects' problem-solving train of thought. If such observations can be supported, it would seem that training for the use of heuristic suggestions must be made powerful enough to permit the incorporation of the heuristic suggestions into the original formulation of plans, conceptions, and decisions, rather than in time-consuming or disruptive reformulations of these processes.

Another group of subjects did, upon reviewing the heuristic prompts, alter the verbal description of their hypotheses. It was discovered, however, that verbal reformulation of hypothesis statements does not always correspond to cognitive reformulation of the appropriate psychological problem space. For example, one subject initially interpreted the cough and shortness of breath symptoms of the case 4 patient as "pneumonia." On recognizing the overly specific character of this early formulation, she changed her hypothesis to "infectious process." However, instead of altering her data collection plan to investigate the possibility of any infectious process (which would have been ruled out on the basis of several low-Cost inquiries) she continued to request information specifically related to the hypothesis of pneumonia.

Finally, there was evidence that some subjects altered both the verbalization of the hypothesis and their conception of the problem through the application of the hypothesis specificity heuristic. For example, in case 3, the symptom of left chest pain initiated a search by one subject for myocardial infarction or angina. On reviewing the hypotheses for undue specificity the subject instead switched his questioning to general cardiovascular functioning, found no significant positive findings and quickly turned his attention to more probable causes of the pain.

Because of the initial variability among subjects on the Scope variable and the observation that only about 50% of subjects even minimally applied the hypothesis specificity heuristic to their early formulations, significant treatment group differences on the Scope variable could not be demonstrated. For specific subjects, however, attention to the Scope of hypotheses did appear to take place and to guide subsequent inquiry. In general, it may be speculated that extensive practice in formulation and evaluation of hypotheses is necessary before the process can be integrated into diagnostic problem solving without interfering with the main line of diagnostic reasoning.

Discussion of Critical Findings Results

Among the five experimental heuristics there were none especially directed toward increasing or decreasing the number of Critical Findings elicited. Instead, each of the experimental heuristics was intended to bring about the more effective use of information as opposed to more thorough accumulation of information. It was not known whether increased effectiveness in the use of information would affect the number of Critical Findings elicited. Although the difference between group means was not statistically significant, the trends suggested the possibility that the effect of the heuristics may have been to reduce in number the Critical Findings elicited with no reduction in diagnostic Accuracy. The correlations between Accuracy of diagnosis and number of Critical Findings elicited, however, indicated a positive relationship between these two variables. The relationship between Accuracy and Critical Findings appears to be partially a function of the kind of data which were included among the Critical Findings. Critical Findings, by definition, have high diagnosticity for the actual pathology of the patient. Prominent among the Critical Findings were the highly specialized tests used primarily as confirmations of presumptive diagnoses. For example, in case 1, a subject who had arrived at a presumptive diagnosis of multiple myeloma on the basis of the patient history, physical exam, and routine urinalysis might confirm his diagnosis with a test for Bence-Jones proteins in the urine. The positive laboratory report for presence of Bence-Jones protein would be considered by experienced physicians to be confirmatory evidence of multiple myeloma. Subjects who had already confirmed the diagnosis of multiple myeloma with the test for Bence-Jones protein, however, were able to inflate their Critical Findings score by reconfirming the diagnosis with additional exotic but unnecessary tests including protein electrophoresis, bone marrow biopsies or aspirations, and metastatic bone surveys. Attempts were made to evaluate the point at which thoroughness of search for new information may have turned to needless reconfirmation of established findings. Such procedures could not reliably be carried out, however, since the data collection procedures did not include measures of certainty of the diagnosis.

The correlational analysis in Table 17 indicates that the number of Critical Findings appears for most subjects to be a function of the thoroughness their questions. The more thorough the collection of information, the more likely the subject was to elicit Critical Findings as well as Moderately Important Findings and Noncontributory Findings. In order for the heuristics to produce greater efficiency, the "dross" rate should be reduced for those subjects exposed to the heuristics; that is, the ratio of Critical Findings to Noncontributory Findings would be greater for subjects trained and prompted to use the experimental heuristics (group 1) than for subjects trained and prompted to use their idiosyncratic heuristics. In fact, the ratio of Critical Findings to Noncontributory Findings for the four treatment groups suggests such a possibility. The proportion of Critical Findings to Critical plus Noncontributory Findings, expressed as a percentage, is presented in Table 22 for each of the four treatment groups.

Table 22

Proportion of Critical Findings to Critical Plus Noncontributory Findings by Treatment Groups

Treatment Group	n	Proportion	S.D. of Proportion
1	8	.340	.018
2	8	.281	.013
3	8	.210	.062
4	8	.248	.055

An analysis of variance failed to demonstrate statistically significant differences between treatment groups ($F_{3,23} = 1.49$, $p < .25$)

Discussion of Cost Results

As with differences on other variables, group means on the Cost variable favored the use of the experimental heuristics. In terms of equivalent dollar values, the Cost differential among the groups was quite large. The mean diagnostic Cost of group 1 subjects, using the experimental heuristics, was approximately half of the mean diagnostic Cost incurred by group 3 subjects, using their idiosyncratic heuristics. Group differences of this magnitude were not, however, statistically significant because of the extreme variability among subjects within groups. The standard deviation on the Cost measure ($215) was approximately four times the anticipated amount of dispersion, and positive skew was substantial.

Observation of subjects revealed at least three different reasons underlying high Costs. One source might be called compulsiveness or inability to separate the important information from the unimportant. Subjects in this category typically elicited a complete history and physical examination, noting any piece of data which might remotely resemble a clue. Each of these marginal and perhaps unreliable findings typically was thoroughly followed up with additional costly procedures. For example, the woman in case 4, when asked about her history of previous surgery, reported that she had a varicose vein stripped in her right leg at age 15. On further probing she explained that the surgery was done for cosmetic reasons and because she was told that it might give her problems later. This finding, in the light of other historical and physical evidence, appeared to most students to be what was commonly referred to as a "red herring." Some subjects, however, despite their own frequent admission that the finding was probably noncontributory, felt compelled to follow up with expensive and risky procedures including arteriography in order to rule out the remote possibility of thromboembolic disease. On finding negative results to their costly procedures, the response of the subjects was usually, "I thought so." This kind of student was fully aware of the high cost and low probability of payoff of his exotic procedures but appeared to disregard these factors in his decision making. In post-problem interviews with these subjects the explanation usually given was that good medicine demands that all possibilities be checked even at some slight risk to the patient. Financial expenses were generally dismissed with references to third-party payment plans.

A second type of student incurred excessive costs because of what appeared to be either inefficient problem-solving skills or inadequate medical knowledge. In short, this kind of student may have had sufficient information on hand to make the diagnosis but failed to recognize the diagnostic significance of his data. Such students simply needed to collect additional information before eliciting evidence which they could correctly interpret in order to secure the diagnosis. Interviews with this kind of student revealed either that they simply did not know the diagnostic significance of an important finding or that they failed to put together the clues which in retrospect seemed obvious to them.

A third kind of student was in some respects a combination of the previously mentioned two. Such subjects (1) failed to interpret data correctly and (2) displayed disregard for diagnostic costs. Typically this kind of student became lost in the problem and exhausted reasonable hypotheses. Rather than admit his deficiencies of clinical skill he tacitly implied that he had deficiencies of data. Subsequent data search often took the form of an unsystematic search for unlikely diagnoses through exotic procedures. Such students were fortunately represented by only three subjects in the present study but their performance is worth mentioning because of the striking similarity between them, and because behavior patterns of this type would seem to warrant detection and counseling of some kind in the interest of future patient care. These subjects all reached a point in the solution of at least one problem in which each of their seriously entertained hypotheses had been ruled out, either correctly or incorrectly. Other subjects, when they found themselves in a similar position, reviewed their data base, made a few more attempts, and as the Initial Instructions suggested, they "referred" the case to a specialist. The three subjects in question were highly reluctant to refer. One of the subjects subjected the patient in case 4, a young woman in extreme distress, to all of the following expensive, risky, and uncomfortable laboratory procedures: Intravenous pyelogram, renal arteriogram, upper gastrointestinal series, cholecystogram, lung scan, bone marrow aspiration, retrograde pyelogram, lumbar puncture, renal biopsy, muscle biopsy, nasogastic tubing for analysis of stomach blood, and a barium enema!

The procedures, listed above in the order requested by the subject, attest to the undisciplined hunt for pathognomonic clues of unsupported hypotheses. While the performance cited above was the most bizarre example encountered, the other two subjects of this type were remarkably similar in their attitude and approach to the patient. Remarks by the subjects indicated that they were unwilling to admit that the problem was beyond their competency ("I think I've got it now." "Now it's falling into place." "I think I should check out a few more things before referring." Etc.) and callous disregard of the patient's condition ("This may kill her, but I'd like to have a renal biopsy.")

In summarizing the experimental performance of the Cost variable, two salient points deserve mention. First, the variability in performance on the Cost variable was so extreme that on this basis alone the factors of expense, discomfort, and risk in diagnostic settings may deserve the attention of those charged with clinical education. Second, the evaluation of diagnostic Costs appears on a subjective level to be a sensitive indicator of diagnostic performance and of various kinds of problem-solving, medical content, and attitudinal deficiencies.

Discussion of Accuracy Results

The pattern of means among the treatment groups on the Accuracy measure suggests that subjects trained and prompted to use the experimental heuristics were more accurate in their diagnoses than subjects trained and prompted to use their idiosyncratic heuristics. The statistical significance level of $p < .07$ for differences among treatment group means is encouraging, but it should be recalled that because each of the four dependent variables was analyzed independently, the probability of reaching statistical significance on the test of at least one variable was increased by four.

Bearing in mind the result that the null hypothesis was not rejected for the Accuracy measure, it is possible to speculate on the pattern of group means obtained. The mean scores for groups 2 and 4, the unprompted groups, are separated by only 1.09 points. It was the experimenter's distinct impression that although the list of experimental heuristics was available to the group 2 subjects, seldom was the list ever used during the posttest problem solving. Thus, for group 2 subjects, the effect of the heuristic training was probably residual; perhaps interpreted as a stronger admonition than that given in the Initial Instructions to All Subjects (see Appendix A) to think carefully and avoid unnecessary procedures. Under these conditions it is not surprising that group 2 subjects performed approximately equally to the untrained, unprompted subjects in group 4.

Group 2 and group 4 subjects may be considered as a control group in the comparison of subjects trained and prompted to maximize use either of the experimental or of the idiosyncratic heuristics. Viewed in this way, the group means suggest that systematic incorporation of the experimental heuristics might be facilitory to diagnostic Accuracy, while systematic use of the idiosyncratic heuristics might be detrimental in comparison to the unprompted controls. Why should more systematic use of one's own heuristics impair the Accuracy of diagnosis?

If the mean Accuracy scores reflect treatment effects, the particular pattern of means is consistent with the hypothesis that the content of heuristic suggestions is crucial, and that systematic prompting to use heuristics will produce different results depending on the quality of the heuristics prompted.

One might ask which of the idiosyncratic heuristics might have been detrimental to Accuracy. This question is difficult to answer since 32 different heuristics were generated and used by the 8 group 3 subjects. It is interesting, however, that for 6 of the 8 group 3 subjects a heuristic stressing thorough collection of data was mentioned. Further, when asked periodically to select a heuristic which fit the present situation, this heuristic was called upon a disproportionate part of the time--almost to the exclusion of other heuristics.

Consequently, for six of eight subjects in group 3, compulsive thoroughness was virtually the only heuristic used. Subjects employing the thoroughness heuristic repeatedly had a mean Accuracy score 6.3 points lower than those not employing this heuristic and a Thoroughness score 5.6 points greater than the remaining members of group 3. Standard deviations for Accuracy and Thoroughness were 6.7 and 14.4 respectively.

The explanation of excessive thoroughness by itself does not convincingly explain the poorer Accuracy performance of group 3 subjects, however, since a nonsignificant relationship between Accuracy and Thoroughness for the entire sample was reported in Table 16. Further, the relationship between Thoroughness and Accuracy may be a complex one.

Drawing upon information processing theory and empirical studies of cognitive complexity by Schroder et al. (1967), one might predict that compulsive thoroughness would lead to information overload and poorer solutions. Unfortunately, the design of this study provided a less than optimal setting for the test of the information overload hypothesis. In the interests of task validity, subjects were permitted to gather data under conditions similar to their normal routine, which included the taking of notes. The notes taken by many of the very thorough students were, unlike the notes taken by the typical experienced physician, extremely detailed. Thus, these subjects provided themselves with an external rote memory and obviated the need for immediate information recoding and chunking. In order to test the information overload hypothesis more adequately, restrictions on the use of external memory aids would have to be incorporated into the design.

Conjectures on Heuristic Processes and the Content of Medicine

There was an additional piece of anecdotal evidence regarding the relationship between compulsive thoroughness of data search and the Accuracy of the diagnosis noted with respect to case 2, the Training Case. The correct diagnosis in case 2 was acute infectious mononucleosis and an underlying chronic form of blood disease, hereditary spherocytosis. Although systematic data were not collected on case 2, it was noted by the experimenter that subjects who tended to be compulsively thorough would reach the diagnosis of infectious mononucleosis and would usually conclude their diagnostic efforts at this point. Those students who had arrived at their diagnosis of infectious mononucleosis by a more direct route usually extended their investigations and were able to arrive at the more significant but less obvious problem of spherocytosis. Both the thorough subjects and the nonthorough subjects had usually turned up the finding of anemia--the one finding which could not be subsumed under the mononucleosis diagnosis. It seemed as though those students who had spent an extensive period of time arriving at their diagnosis were more willing to let this one loose end remain unresolved. The nonthorough subjects, having spent a considerably shorter period of time on the problem to that point, were willing to investigate further the underlying cause of the anemia. Thus, one might speculate that even if the information overload hypothesis does not apply to diagnostic problem solving among advanced medical students, compulsively thorough diagnoses may be construed as a misallocation of diagnostic energies.

During the planning of this study, the observation of subjects involved in problem solving, and the subsequent analysis, the writer has searched for a reasonable explanation for the strategies of problem solving employed and for better explanations of problem-solving behavior. It is clear at the end of these many months of conjecture and observation that the strategies and heuristics used in solving diagnostic problems are more complex, situation-specific, and content-specific than previously imagined.

An example of the situation-specific nature of the heuristics was provided in the early pilot testing of the simulated cases. In the original instructions, the experimenter played the role of a patient during the medical history portion of the work-up. The physicians were unable to follow instructions which, in essence, asked the doctor to ignore any possible sensitivities to the patient. If the doctor immediately suspected alcoholism, for example, he was permitted to pursue it directly without the need to first establish rapport. The physician subjects felt unnatural ignoring some of the strategies they typically used when the patient was in the room with them.

In order to eliminate the personality and interpersonal aspects of the doctor-patient relationship which were beyond the scope of this study, a second setting was evolved in which the subject was told to think of the experimenter as a computer terminal which could retrieve any piece of information requested about the patient. Subjects operating in this mode felt completely unconstrained in the sequence of information collected and felt no compunction about ordering exotic, expensive, and dangerous diagnostic procedures since the information had presumably been obtained and filed previously. Subjects reported feeling uneasy with this style since their diagnostic plan became an unsystematic search for pathognonomic laboratory tests intended to rule out vague hunches.

Instructions placing the setting of the diagnostic work-up in an office practice resulted in clearly different behavior from a diagnostic work-up in a university hospital. Emergency Room diagnostic strategies were different from those applied to admitted patients. In many ways, the decision about what information to collect was based upon heuristics peculiar to the diagnostic setting. The final setting decided upon for testing the subjects was an out-patient clinic in which all of the facilities of the hospital would be available, if needed, but one in which problem-solving heuristics were not necessarily tied to hospital routines. This setting, in which the experimenter acted as an intermediary third person reporting whatever information was requested, eliminated the variable of the doctor-patient relationship and forced all historical inquiries into a closed-ended type.

The situation specificity of the heuristics became obvious in the prompting of subjects to use the heuristics. It appeared that the number of heuristics used by a subject during a single diagnostic work-up was very much larger than five, and the heuristics used were almost exclusively of a conditional nature. Strategies were selected depending upon what information was currently available, the perceived health and comfort conditions of the patient, and the particular hypotheses under consideration at the time.

The Relative Importance of Knowledge and Strategy

The relative importance of knowledge of the content of medicine as opposed to skill in problem-solving process is an issue that has been debated in medical education for some time. The experimenter had the distinct impression during the problem-solving sessions that heuristic processes were of secondary importance to the students' knowledge of the content of medicine required to solve the diagnostic cases. For example, in case 4 a finding of red cell casts in the urine was almost invariably elicited since it was reported to all subjects requesting a routine urinalysis. The significance of this finding, a nearly pathognomonic sign of glomerulonephritis, was missed by approximately 40% of all subjects. It was clear that regardless of the heuristic processes employed, deficiencies of knowledge of the correct interpretation of this finding were of the utmost importance in determining both the Cost and Accuracy of the work-up for the case. Further, it appears in retrospect, that the knowledge of subjects was likely to vary widely across problems. Some subjects performing in an outstanding manner on one problem performed very poorly on another, and the differences appeared from the debriefing to be often due to lack of specific information or misinformation about the significance of the findings elicited.

Great gaps in knowledge are, perhaps, to be expected among students beginning their fourth year of medical school. The previous experience of these students consisted of two years of training in the basic sciences and 9 to 12 months of clinical experience. The types of clinical experience during the third year of medical school vary because of scheduling problems and specialty choices of the students. It cannot be said that the previous exposure of students to the medical content required by the cases in the study was equivalent.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The ability to reach accurate diagnostic conclusions while treating patients humanely is a major goal of medical training. Medical schools have typically assumed that better diagnosis was to be achieved through the Baconian ideal of thorough and impartial gathering of facts which are later objectively interpreted and evaluated.

Systematic observation of competent practicing physicians, however, has led to the conclusion that the process of diagnosis is one in which hypotheses are continually advanced, tested, modified, ruled out, or confirmed. Physicians collect medical case data almost exclusively for the purposes of generating hypotheses and aggregating evidence in their favor.

There are obvious dangers in allowing hypotheses and conjectures to influence data collection and interpretation including premature closure, selective information gathering, and biased interpretation. Conversely, there is reason to believe that these hypotheses may serve an indispensible function even in the earliest stages of the work-up. The formation of hypotheses appears to direct the search for information. In addition to the greater economy of focused rather than thorough data collection, hypotheses appear to function as the organizing principles for the storage and recall of information in memory.

This study has taken the position that the dangers of hypothesis-guided diagnostic inquiry should not be countered by struggles to eliminate early hypotheses, but instead by training in diagnostic heuristics which might help diagnosticians to generate more adequate hypotheses and to test their hypotheses more effectively.

A set of five experimental heuristics was derived from analysis of the reported and observed errors of diagnostic reasoning committed by medical students. Thirty-two advanced medical students attending two Michigan medical schools were selected as experimental subjects.

In order to test the thesis of this study and to obtain evidence of the effects of various kinds of heuristic content and usage, the students were presented with a series of medical cases which they were to diagnose. Each student was assigned to one of the four following conditions of heuristic training and prompting:

Group 1: Subjects were given prior training in the use of the experimental heuristics and periodic prompting to employ these heuristics in their problem-solving effort.

Group 2: Subjects were given prior training in the use of the experimental heuristics and were invited to employ the heuristics at their discretion.

Group 3: Subjects were given an orientation in the use of heuristics, were asked to generate a list of personal heuristics they had found useful in diagnosis, and were given periodic prompting to employ their own heuristics.

Group 4: Subjects were given no prior training and only a brief orientation to the use of heuristics and no prompts which might influence their problem solving.

All subjects were asked to solve the diagnostic problems as efficiently and as accurately as possible. Four measures of problem-solving performance were taken for each subject on each diagnostic case. The dependent measures were defined as follows:

<u>Scope of the Early Diagnostic Formulations</u>.--The Scope measure was intended to reflect the degree of

generality or specificity of early hypotheses. This measure was included because overly specific early hypotheses have been reported to be associated with premature narrowing of the diagnostic problem space and subsequent poor problem-solving performance.

Number of Critical Finding Elicited.--The Critical Findings measure was intended to assess the extent to which subjects elicited the particular pieces of information judged most useful in arriving at the correct diagnosis of each case presented. The relationship between obtaining highly diagnostic information and making accurate diagnoses was of particular interest.

Cost of the Diagnostic Work-up.--The Cost measure was defined as an additive function of financial expense, patient discomfort, and risk to patient health inherent in the diagnostic work-up. Medical schools in general have been frequently accused of paying too little attention to these aspects of patient care.

Accuracy of the Diagnosis.--The Accuracy of diagnosis was defined in such a way as to give greater credit for diagnosing the primary problem than for secondary complications or unrelated minor problems, greater credit for increasing specificity of diagnosis, and negative credit for incorrect diagnoses.

The Scope and Critical Findings measures were considered to be process measures which might be related to diagnostic outcomes. The measures of Cost and Accuracy were considered to be diagnostic outcomes of paramount importance.

The contribution of this study is two-fold. First, a set of dependent measures for the quantification of important diagnostic outcomes has been defined and investigated. Those investigations demonstrated that the evaluation of diagnostic Cost and Accuracy performance can be made objectively, but that several cases will probably be required to obtain acceptable coefficients of reliability. Second, the effects of problem-solving heuristics on process and outcome measures have been investigated as well as relationships among many performance variables. The principle findings resulting from these investigations were as follows:

1. There was no acceptable evidence that the heuristic training or prompting effected the performance of subjects on any of the four principal dependent measures.

2. Treatment group differences on the Accuracy measure approached acceptable levels of significance ($p < .07$). On this basis the hypothesis of treatment effects on the Accuracy measure is judged to be worthy of further pursuit.

The trends between treatment groups suggested that the experimental heuristics were more beneficial than the idiosyncratic heuristics.

3. No significant relationship was found between the Scope of Early Diagnostic Formulations and either the Cost or the Accuracy of diagnosis.
4. The number of Critical Findings elicited was positively associated with both higher Cost and greater Accuracy, but no significant relationship was found between Cost and Accuracy.
5. No relationship was found between medical College Admission Test scores administered prior to entering into medical school and measures of diagnostic Cost or Accuracy.

The general hypothesis of this study, that heuristic training might improve the problem-solving performance of advanced medical students functioning in a hypothesis-guided mode, has not been supported by the findings. Trends with respect to the group means on the diagnostic Accuracy variable, however, are encouraging evidence in favor of the hypothesis. The findings may also be interpreted as failing to support some of the previously untested assumptions of current pedigogical practice in medical education. Specifically, the results of this study indicate that greater thoroughness of the history and physical examination is associated with greater diagnostic Cost but is not associated with greater diagnostic Accuracy.

Critical Review of the Procedures

The popular view of the scientist as a person who advances the frontiers of knowledge one well-placed step at a time is not entirely accurate. As in other endeavors, scientific research has room for the demands of different problems and the styles of different researchers. So from time to time an advance scouting party will, at some risk, set out to seek an answer with an optimism unsupported by great probability of success. The present study was exploratory in the sense that findings were sought in areas many steps removed from the ground covered in the Review of Literature. The camp from which these investigations began was defined principally by Elstein and Shulman and their work with the normative problem-solving behavior of experienced physicians. There are many intermediate steps between this base and the training of advanced medical students in problem-solving skills. Some of these steps turned out to be more difficult to negotiate than expected, others were taken in stride. The intent of this section is to review the procedures of the research so that others who follow may be better prepared to foresee the pitfalls and snares along the way.

The first experimental difficulty resulted from the extreme variability among subjects on the dependent measures. Much of the dispersion appeared to be a function of the limited and specialized clinical exposure of the students. Those subjects who had spent the previous six months working in general surgery, orthopedics, and radiology were at a distinct disadvantage to those subjects who had rotated through internal medicine, hematology, or gastroenterology clerkships. Greater experimental precision could probably be obtained through a process of matching students on the basis of previous relevant clinical experience and randomly assigning cohorts to treatments.

A second major difficulty lay in the definition of the hypothesis-guided approach. Subjects were defined as operating in the hypothesis-guided mode because they were required by the design to verbalize hypotheses at regular intervals and were instructed to ask only those questions which they believed would be helpful in arriving at the diagnosis. Observation of subjects' approaches to solving the diagnostic cases, however, indicated that for a significant proportion of the subjects, the hypotheses, though verbalized, were not guiding the data-acquisition process. Some subjects performed naturally in a hypothesis-guided mode, others became reluctantly hypothesis guided, and a few persisted in a step-wise approach.

A third difficulty lay in the training for the use of heuristics. The literature search pertaining to heuristic training had left unclear whether the value of heuristics resided in heuristic instruction which might alter the content learned as well as the reasoning processes applied to content, or only in the reasoning processes themselves. The present study attempted to separate these two possible types of effect by applying heuristics systematically to the analysis of medical content learned previously under nonheuristic instructional methods. The experimental heuristics were, in a sense, tacked on to the medical instruction at the last minute. Based on observation of the performance of subjects, the writer has been led to the conclusion that there was insufficient practice time provided in which the subjects could integrate the heuristics meaningfully into their on-going train of problem-solving thought. Perhaps the only fair test of the value of heuristic reasoning is one in which students are taught how to solve problems through continuous asking and answering of heuristic questions. Such a procedure inevitably leaves confounded the original question of content versus process effects of heuristics, which the present design attempted to separate. The apparent circularity of the problem may result from fundamental misconceptions of the nature of and relationships between the constructs of content and process. It would seem that further work in defining the psychological effects of heuristic usage must first illuminate the nature of any content-process distinction.

For purposes of the practical curricular employment of heuristics the psychological issues of process and content seem less important. The meager existing evidence on the effectiveness of heuristics favors the systematic incorporation of heuristic methods into the teaching of content rather than the subsequent addition of heuristic suggestions.

A fourth design problem was the limited number of cases which could be presented to each subject. The finding of no significant relationship between Accuracy scores on the two posttest cases was disappointing but not an unusual finding. McGuire and Lewy (1966) reported that to reliably assess the problem-solving performance of subjects in the Patient Management Problem (PMP) format, at least 12 lengthy problems would be required. Researchers in Elstein and Shulman's group (1973) found no relationship between the proficiency scores of patient management on the PMP's and the accuracy of diagnoses presumably underlying the physician's management plan. Reports from these and other researchers in medical problem solving seem to point to two conclusions: (1) the adequate assessment of problem-solving skills of medical students requires many simulated cases of extended length and

(2) the skill and knowledge dimensions which are currently lumped together under the rubric of problem solving need to be more carefully delineated.

In review, it would appear that some of the problems encountered in the present study might well yield to refinements of experimental design. Other problems may be at present intractable and will only yield to extensive commitment of resources and ingenuity. Viewing the difficulties encountered in the present study against the background of the preceding research a number of implications for future research and development might be advanced.

Implications for Future Research and Development

The implications for development as well as for research are to be discussed because the thrust of research in medical problem-solving must eventually reach into curricular systems for the training of physicians. First, however, implications for basic psychological research will be examined.

As investigators of cognitive processes have raised their sights from simple anagram and match-stick problems to diagnostic problems in medicine, electronics, and social systems, immense complexities have appeared for which the present theoretical base is inadequate. An adequate theory of medical problem solving will probably require at a minimum the interrelationship of what are presently discrete theories of memory, information processing, and decision making. All of these dimensions are clearly enmeshed in the problem-solving activity of the physician. At another level, distinctions within memory theory, information processing theory, and decision-making theory have frequently turned out to be conceptualizations with narrow empirical evidence of psychological reality. Promising new constructs in problem-solving commonly fail to generalize across problems of different types. Newell and Simon (1972) have explained that many constructs of problem-solving must inevitably be task-specific, but some problems appearing to be equivalent in their task requirements also fail to yield comparable results in many cases. Despite recent theoretical advances, it is difficult to ignore the possibility that the constructs with which we presently build models of problem-solving are at an extremely crude level of development.

Another question for the problem-solving theoretician is the relative importance of knowledge of content versus skill in problem solving. Could or should the process of problem solving be taught in a pure form with expectation of transfer to broad categories of content? Or, should teachers supply heavy doses of content and expect the useful employment of content in problem-solving situations to be a function of the amount of stored knowledge in the head of the problem solver?

The present study has demonstrated the use of two promising outcome measures of diagnostic problem solving. The process measures believed to be associated with successful diagnostic problem solving were not predictive of the measures of success. Within the domain of diagnostic problem solving considerable theoretical and empirical work will be required to define the behaviors that regularly lead to successful problem solving. The definition of such behaviors will most likely represent a significant theoretical advance, either leading to or confirming the existence of useful constructs underlying diagnostic problem-solving performance. In immediate and practical terms the identification of behaviors associated with superior problem solving may provide a key for effective problem-solving training.

The scoring of clinical simulations has also created problems of a theoretical nature since these testing procedures do not lend themselves well to analysis under the classical measurement model. Difficulties begin with the definition of the test item. If we consider each piece of information received as an item which the subject may interpret correctly or incorrectly, we must face the following problems:

1. Subjects may choose their own items.
2. Items vary widely in their score values.
3. Total test length varies as a function of the subject's item-choosing behavior.
4. Items are not mutually independent.
5. Correctness or incorrectness of interpretation of an item is probabilistic.
6. Correctness or incorrectness of an item must often be inferred from the content of subsequent items chosen.

Simulation instruments such as those developed for the present study have in essence traded off some of the advantages of psychometric elegance in favor of greater task validity (Shulman, 1970). The unit of behavior analyzed in the present study was a lengthy sequential activity in which subjects decided what information to collect depending on previous questions asked and the responses elicited. Unlike tests of knowledge where independent items can be considered as sampled units of behavior, the interest in diagnostic problem solving focuses on the entire sequentially dependent, complex pattern of alternate search and interpretation activities leading to a unique diagnostic conclusion. To remove the sequential dependency among items or the free choice of information elicitation would alter the task to the extent that the measures of performance would no longer reflect the processes or outcomes representative of diagnostic problem solving. Consequently, in an effort to provide valid exercises of diagnostic activity, the determination of other psychometric properties has been made more difficult.

The issue of validity of simulations does not rest entirely on the achievement of task validity. As with other evaluation instruments, the question must always be raised, validity with respect to what? In the present study, task validity was considered to be of paramount importance, but interpretations are necessarily limited to performance on problems of a particular type. Specifically, the cases presented represented problems in internal medicine, dealing only with diagnosis of serious and somewhat complicated organic illness presented by new adult patients.

Simulations designed for purposes other than the research questions of this study might be concerned with other kinds of validity. For example, simulations intended to certify the general competence of physicians might require greater attention to ecological validity; assuring that problems are more broadly representative of the kinds of cases typically encountered in a given kind of medical practice.

Medical school curriculum developers can be optimistic about the use of clinical simulations because studies of the present type demonstrate that it is possible to define and measure at least some of the components of clinical judgment which have traditionally been evaluated only by global impressions of students by faculty. While personal assessment must continue to be a crucial part of medical training, the use of well-structured simulated cases holds the promise of providing medical educators and medical students with the means of more reliably defining particular areas of deficiency and strength.

Simulated problems in medical diagnosis and patient management can never completely replace experience with actual patients but can extend the clinical experience of medical students in important ways. First, simulations offer the opportunity to draw diagnostic conclusions and to make management decisions under conditions where live patients are not subjected to the risk of less-than-expert care. Evaluation of clinical competence can be made with reference to criteria set by experts and to normative performance measures based on large samples of medical students tested under standardized conditions.

The face-to-face format of the present study is admittedly a time-consuming and inefficient process.

Advances in computer system technology including natural language capability and time sharing networks among medical schools have, however, provided the means for efficient distribution, presentation, scoring, analysis, and interpretation of student problem-solving performance. Looking further into the future, projected plans for medical satellite communication networks may bring about the possibility of the teaching and evaluation of some clinical skills in remote sites.

Optimism for the promise of clinical simulation must be tempered by caution. As with any new tool, the possibilities of misuse are great. One of the significant deficiencies of such instruments lies in the presently inadequate conceptual models for interpretation of performance scores. The theoreticians have not yet defined the relevant psychological dimensions of medical problem-solving ability. Until the structure of abilities which comprise adequate problem-solving performance are identified, student evaluation based on these performance measures must remain qualified.

The task of securing evidence for the validity of clinical simulation is one of paramount importance. A hopeful note with respect to simulation validity is that new designs for assessing the validity of Patient Management Problems have been formulated (Sedlacek and

Nattress, 1972) and preliminary evidence that PMP's can at least identify the worst of practitioners is somewhat encouraging (Goran, Williamson, & Gonnella, 1973).

A final caveat for users of simulations bears on the relative importance of simulation performance with respect to other dimensions of student performance. The danger exists that skills which can be quantified most readily may come to exert a controlling influence over the teaching aims of medical schools. Diagnostic skills, patient management skills, and knowledge of medical content are presently more objectively quantifiable than interpersonal skills and medical student attitudes toward patients. In 1973 the largest task of medicine has shifted from the chemotherapeutic treatment of acute infectious processes to supportive care in chronic debilitating illnesses for which there is no cure (Glazier, 1973). The goals of medical education must be derived from the rational assessment of patient needs. It is the responsibility of the curriculum developer to see that the available technology is used to further these ends and not to subvert them.

SELECTED BIBLIOGRAPHY

SELECTED BIBLIOGRAPHY

Allal, L. K. Training of second-year medical students in diagnostic problem formulation. Unpublished Ph.D. dissertation, Michigan State University, in progress.

Ashton, M. R. _Heuristic methods in problem solving in ninth-grade algebra_. Ann Arbor, Mich.: University Microfilms, 1962.

Attneave, Fred. _Applications of information theory to psychology_. New York: Holt, Rinehart and Winston, 1959.

Barrows, Howard S., & Bennett, Kara. The diagnostic problem solving skill of the neurologist. _Arch. Neurol._, 1972, _26_, 273-277.

Blain, A., III (Chairman). A report of the committee on relative value study. Lansing, Michigan, 1971.

Boring, E. G. _A history of experimental psychology_. New York: Appleton-Century-Crofts, 1950.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. _A study of thinking_. New York: Wiley, 1956.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), _Handbook of research on teaching_. Chicago: Rand McNally, 1963.

Chamberlin, T. C. The method of multiple working hypothesis. Reprinted in _Science_, May 7, 1965, _148_ (3671).

Covington, M. V.; Crutchfield, R. S.; Davies, L.; & Olton, R. M. _The productive thinking program_. Columbus, Ohio: Merrill, 1972.

Cronbach, L. B., & Rajaratnam, N., et al. Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 1963, *16*, 137-163.

de Groot, Adriaan D. *Thought and choice in chess*. The Hague: Mouton, 1965.

Dewey, John. *How we think*. Boston: D. C. Heath and Co., 1933.

Ebel, R. L. Estimation of the reliability of ratings. In W. A. Mehrens & R. L. Ebel (Eds.) *Principles of educational and psychological measurement*. Chicago: Rand McNally, 1967.

Edwards, W. The theory of decision making. *Psychological Bulletin*, 1954, *51*, 380-418.

Einhorn, H. J. The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 1970, *73*, 221-230.

Elstein, Arthur S.; Kagan, Norman; Shulman, Lee S.; Jason, Hilliard; & Loupe, Michael J. Methods and theory in the study of medical inquiry. *J. Med. Education*, 1972, *47*, 85-92.

Elstein, A. S. An hypothesis testing model for the study of medical judgment. Paper presented to the Fifth Annual Conference on Human Judgment, University of Colorado, Boulder, Colorado, March, 1972.

Elstein, A. S. Personal communication at American Educational Research Association meeting, New Orleans, 1973.

Ernst, G., & Newell, A. *GPS, a case study in generality and problem solving*. New York: Academic Press, 1969.

Glazier, W. H. "The task of medicine. *Scientific American*, 1973, *228*, 13-18.

Goran, M. J., Williamson, J. W., & Gonnella, J. S. The validity of patient management problems. *J. Medical Education*, 1973, *48*, 171-177.

Harrison's *Principles of internal medicine*. (6th ed.) New York: McGraw-Hill, 1970.

Harvey, A. M., & Bordley, J. Differential diagnosis. Philadelphia: Saunders, 1970.

Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 57, 116-131.

Holt, J. C. How children fail. New York: Pitman, 1964.

Hoyt, Cyril J. Test reliability estimated by analysis of variance. In W. A. Mehrens & R. L. Ebel (Eds.), Principles of educational and psychological measurement. Chicago: Rand McNally, 1967.

Hunt, E. B. Concept learning: An information processing problem. New York: Wiley, 1962.

Hunt, E. B. Selection and reception conditions in grammar and concept learning. J. Verbal Learning and Verbal Behavior, 1965, 4, 125-211.

Jacquez, John A. The diagnostic process: Problems and perspectives. In the The diagnostic process. The proceedings of a conference sponsored by the Biomedical Data Processing Training Program of the University of Michigan, May 9-11, 1963.

Jensen, A. R., & Rohwer, W. D., Jr. Syntactical mediation of serial and paired-associate learning as a function of age. Child Development, 1965, 36, 601-608.

Johnson, Donald M. Systematic introduction to the psychology of thinking. New York: Harper and Row, 1972.

Kleinmuntz, B. The processing of clinical information by man and machine. In B. Kelinmuntz (Ed.), Formal representation of human judgment. New York: Wiley, 1968.

Larsen, C. M. The heuristic standpoint in the teaching of elementary calculus. Ann Arbor: University Microfilms, 1960.

Lee, Wayne. Decision theory and human behavior. New York: J. Wiley and Sons, 1971.

Lorayne, Harry. How to develop a super-power memory. New York: Frederick Fell, Inc., 1957.

Loupe, M. J. The training of problem solving and inquiry. Unpublished Ph.D. dissertation, Michigan State University, 1969.

Lusted, Lee B. Decision-making studies in patient management. New England Journal of Medicine, 1971, 284, 416-424.

Maier, N. R. F. Mechanization in problem solving: The effects of einstellung. Psychological Monographs, 1942, 54 (6), 1-15.

McGuire, C., & Lewy, A. A study of alternate approaches in estimating the reliability of unconventional tests. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Ill., Feb., 1966.

McGuire, C. A. & Solomon, L. Clinical simulations. 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. Plans and the structure of behavior. New York: Holt, Rinehart, and Winston, 1960.

Platt, J. R. Strong inference. Science, Oct. 16, 1964, 146 (3642).

Polya, George. How to solve it. (2nd ed.) Garden City, N.Y.: Doubleday, Anchor Books, 1957.

Polya, G. On the curriculum for prospective mathematics teachers. American Mathematical Monthly, Feb. 1958, 65, 101-104.

Popper, Karl R. Conjectures and refutations: The growth of scientific knowledge. New York: Harper & Row, 1965.

Reese, H. W. Imagery in paired-associate learning in children. Journal of Experimental Child Psychology, 1965, 2, 290-296.

Reitman, W. R. Cognition and thought. New York: Wiley, 1965.

Rubel, R. A. Decision analysis and medical diagnosis and treatment. Ann Arbor, Mich.: University Microfilms, 1970.

Schroder, Harold M., Driver, Michael J., & Streufert, Siegfried. Human information processing: Individuals and groups functioning in complex social situations. New York: Holt, Rinehart, and Winston, Inc., 1967.

Schwartz, Steven H., & Simon, Roger I. Differences in the organization of medical knowledge among physicians, residents, and students. Journal of Structural Learning, 1972, 3, 23-26.

Schwartz, Steven H., & Simon, Roger I. Information processing and decision making in medical diagnosis. Paper presented at symposium on Health Science and the Systems Approach, Wayne State University, Detroit, March, 1970.

Sedlacek, William E., & Nattress, Leroy W. A technique for determining the validity of patient management problems. Journal of Medical Education, April, 1972, 47, 263-266.

Shannon, C. E., & Weaver, W. The mathematical theory of communication. Urbana: University of Illinois Press, 1949.

Shulman, L. S., Loupe, M. J., & Piper, R. M. Studies of the inquiry process. Educational Publication Services, College of Education, Michigan State University, 1968.

Shulman, L. S. Reconstruction of educational research. Review of Educational Research, June, 1970, 40, 371-396.

Simon, Herbert A., & Newell, Allen. Human problem solving. Englewood Cliffs, N.J.: Prentice Hall, 1972.

Smith, Q. C. A comparison of a heuristic and a traditional method of teaching a preparatory course in mathematics to college freshman and sophomores. Ann Arbor: University Microfilms, 1967.

Sprafka, Sarah. The impact of hypothesis generation and verbalization on certain aspects of medical problem solving. Unpublished Ph.D. dissertation, Michigan State University, 1973.

Von Neumann, J., & Morgenstern, O. Theory of games and economic behavior. (3rd. ed.) Princeton, N.J.: Princeton University, 1953.

Wason, P. C., & Johnson-Laird, P. N. (Eds.) Thinking and reasoning. Baltimore: Penguin Books, Inc., 1968.

Williamson, J. W., Alexander, M., & Miller, G. E. Continuing education and patient care research. J.A.M.A., Sept. 18, 1967, 201, 938-942.

Wilson, J. W. Generability of heuristics as an instructional variable. Ann Arbor: University Microfilms, 1967.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wortman, P. M. Medical diagnosis: An information-processing approach. Computers in Biomedical Research. Academic Press, 1972.

Wortman, P. M., & Kleinmuntz, B. The role of information processing models of problem solving. Unpublished article, Duke University, 1972.

APPENDICES

APPENDIX A

INSTRUCTIONS, TRAINING NOTES, AND ADMINISTRATIVE FORMS

APPENDIX A

MICHIGAN STATE UNIVERSITY EAST LANSING • MICHIGAN 48823

OFFICE OF MEDICAL EDUCATION RESEARCH AND DEVELOPMENT
FEE HALL EAST • TELEPHONE (517) 353-7791

This letter is intended to give you some information about the study of medical diagnosis in which you have agreed to participate. The main purposes of the research are to: (1) characterize the mental operations of advanced medical students as they attempt to solve diagnostic problems, (2) attempt to identify strategies that appear to be helpful in attacking diagnostic problems so that useful strategies can be taught to future medical students, and (3) to field test several new ways of assessing complex problem solving skills. There is no intention to evaluate the clinical competence of individual students.

In the study you will be given a series of diagnostic problems to solve. The format will be as follows: You will be given a chief complaint and a few routine pieces of information about a patient. Your task will be to ask for whatever further information you feel is desirable to reach a diagnosis. The information you request will be supplied by the experimenter.

Following the diagnostic cases you will be given a debriefing so that the experience will be instructive to you as well as providing data for us. The problems are similar to those you may encounter on part 3 of the National Board exams and might be considered as an opportunity to practice on this kind of task.

Since several of your fellow students will be participating in the study, I will ask you to refrain from discussing any aspect of it -- experimental procedures, clinical findings or possible diagnoses -- with your peers until all of the data is in.

The time requirements and honoraria will differ among participants. For your participation the honorarium (contingent upon completion of the entire series of problems) will be $_____. The time and place reserved for your participation are written in the box below. Please tear off this "appointment slip" and keep it as a reminder. Note the number to call if any changes are necessary.

Many thanks for your help.

Sincerely,

Mike Gordon

ks

Diagnosis Study Appointment

Session 1: ____________________

Session 2: ____________________

Place: _______________________

IF UNABLE TO KEEP APPOINTMENT, PLEASE CALL
(517) 353-9656 COLLECT AS SOON AS POSSIBLE.

Mike Gordon

INITIAL INSTRUCTIONS FOR ALL SUBJECTS

Imagine yourself to be an internist in the outpatient department of a 300 bed community hospital. A relatively large proportion of the patients you see in this setting come to the hospital because they have no regular doctor or because they have become ill while away from home.

Your task will be to diagnose the problems of a few such patients by asking for whatever information you believe would be helpful in reaching the diagnosis. I will play the part of a third party who will get the patient history, physical exam results or any lab tests that you request and report them to you immediately. You may consider the information you get to be as reliable as it would be under actual circumstances.

Your goal is to come as close to the definitive diagnosis as possible but without subjecting your patient to unnecessary expense, inconvenience discomfort or risk to his health. In short, get the information you need to make the diagnosis, but be efficient. If you get to a point where you have begun to "spin your wheels" you will probably be better off calling in a consultant (and terminating the problem) than running a series of exotic tests with small chance of yielding the diagnosis.

Since we are interested in diagnosis only on these problems you may eliminate any questions intended only to establish rapport or to make treatment or management decisions.

I will interrupt the question and answer format periodically to ask you what diagnostic formulations or hypotheses you may be entertaining. In making these hypothesis statements feel free to be as general or as specific as you wish.

Here is some basic information to get you started on your first patient. You may take notes on the information you collect if you wish.

Any questions before we begin?

Heuristics Orientation Notes

Heuristic Orientation for All Subjects

Now that you have completed the first case let's look back over what you did. One obvious thing is that you tailored the questions you asked to suit the likely kinds of illness in a person of this age, sex and with these particular kinds of symptoms. In other word what you did was not completely routine.

All the way through these kinds of problems, a doctor has to make decisions about what kind of information is worth collecting, how much information can be put together. Part of clinical experience is getting the feel of making those little decisions. When I say "getting the feel" I am really talking about the fact that each doctor develops a set of decision rules. Not hard and fast rules--but rules of thumb. Some of these rules of thumb get passed along in medical training but many remain implicit and are relearned by every medical student through his own experience. The purpose of this study is to focus on these rules of thumb and determine what kind of part they play in diagnosis.

Heuristic Orientation Supplement for Treatment Groups 1 and 2

We have gone to some lengths studying doctors from university hospitals, private practice and salaried group practice. We have compared average doctors with doctors considered to be expert diagnosticians by their colleagues trying to discover things about their reasoning processes. By working with these doctors and putting together pieces from psychological theories and experiments in problem solving, we believe we have deduced several of those implicit rules of thumb used by good diagnosticians. Now we would like to make these same rules of thumb explicit, to orient medical students to them and see what happens when they are used.

The rules are general enough to apply to virtually any non-trivial diagnostic problem. The list of rules is small because some rules are so widely used and understood that it would be a waste of time to teach them to advanced students. For example, one rule is that you assume that patients middle-aged or younger have only 1 disease which will account for all the signs and symptoms. That rule hardly needs to be taught. Other rules of thumb are peculiar to only a few doctors and including these would make the list too long to remember. So, we have a list of just five rules. Here they are.

HEURISTICS

1. Each piece of information requested by the problem solver should be related to a plan of attack for solving the problem. There should be a plan and a well defined purpose behind every question asked.

    Rationale: Problems can be divided into sub-problems that limit the search. Sub-problems usually have a logical order which, if systematically followed tends to reduce confusion and increase the efficiency of search.

    Examples:
    a. Plan to find whether the chief complaint is the real reason for the visit.
    b. Plan to determine whether the problem is acute or chronic.
    c. Plan to get enough general background on the case to focus in on some small number of likely problems.

2. No hypothesis should be more specific or more general than the evidence on hand justifies.

    Rationale: Hypotheses are used to organize the information collected, and to distinguish between possible problems. If hypotheses are too general the interpretive value of some pieces of information is often overlooked. If hypotheses are too specific interpretive information can be similarly overlooked or appropriate questions not asked at all.

    Examples:
    a. In the case of an anemic negro boy, a Dr. asked for a smear to check his too specific hypothesis of sickle cell anemia. Although other highly diagnostic abnormalities were visible on the slide, they were not seen.
    b. Williamson study demonstrated that lab values routinely taken, even when results are grossly abnormal are not processed unless the Dr. had in mind a hypothesis to which the lab study was relevant.
    c. Failure to focus problem - Some Drs. tend to overestimate the ambiguity of data, leaving hypotheses at such a general level that even after extensive data has been collected, they are unwilling to commence more detailed investigation.

3. There should always be at least 2 to 3 competing hypotheses under consideration at a particular time. Each piece of information should be evaluated with respect to all hypotheses presently under consideration.

    Rationale: Lack of competitors leads to a confirmatory set; the seeking of information relevant to only one problem, and selective perception, selective forgetting of negative data, and to biased interpretation of findings.

    Examples: In M.S. case, Dr. asked about "any eye trouble?" Patient answered, "Not lately." He later elicited a Babinsky but failed to see it. During the work-up he had convinced himself that the problem was hysteria.

4. Whenever a ne
collected (pa
questions aske
categorize the
tending to dis

Rationale

5. When high cost
to confirm a fa
of lower cost
possibilities
the probability

Rationale:

Example:

HEURISTICS - 2

4. Whenever a new or revised hypothesis emerges, the information previously collected (particularly the information from the middle of the sequence of questions asked) should be reviewed. The problem solver should attempt to categorize the previously elicited findings as either tending to confirm or tending to disconfirm his new hypothesis.

Rationale: Research has demonstrated that selective forgetting takes place for informative data not explainable within hypotheses on hand. Even if the hypothesis is later entertained, data collected prior to the generation of the hypothesis and consistent with it is infrequently used to support the hypothesis.

5. When high cost(expensive, uncomfortable or risky) procedures are being considered to confirm a favored hypothesis, the problem solver should consider the possibility of lower cost procedures which might instead <u>rule</u> <u>out</u> one or more diagnostic possibilities in order to make the high cost procedure unnecessary or to increase the probability that the high cost procedure will yeild the definitive diagnosis.

Rationale: Confirmatory sets often preclude the use of more simple procedures which can lead to rapid elimination of alternatives. While such procedures (negative inference) may have equivalent power to reduce uncertainty, they tend to be under-utilized for psychological reasons and by habit.

Example: A Dr. held a high priority hypothesis (on the prior case) of a type of ulcer and a lower priority hypothesis of a hematologic problem. Rather than ruling out his hematologic hypothesis with some simple blood tests, he first proceded to order an upper and lower G.I. series.

Heuristi

Wh
of thumb
automati

1.

2.

3.

4.

Heuristic Orientation Supplement for Treatment Group 3

What I would like you to do is to see if you can think about what rules of thumb you use in diagnosis. That may be hard to do since they become so automatic, but here are a few ways to get you started.

1. Think of the silly kinds of errors you may have made in the past and told yourself you'd never do that again. State a rule that would help you never to do that again.
2. Think of the admonishments that you continually hear from faculty and believe to be helpful.
3. Think of a list of do's and don'ts of diagnosis that you would endorse.
4. Think of yourself as trying to diagnose a case with a supervisor who is asking you leading questions to aid your thinking. Take some of the questions he is asking and put them into the form of rules.

APPENDIX B

PROBLEM-SOLVING CASE DATA

A
to your
by a fr
have no
informa

oc
he
we
ag
te
cl

APPENDIX B

CASE 1

A male caucasian patient appearing to be in his early 20's comes to your hospital outpatient department at 10:30 Monday morning accompanied by a friend. He appears weak, fatigued and underweight. He claims to have no regular doctor. The nurse has collected the following routine information:

occupation	student
height	5'10"
weight	145
age	22
temperature	99
chief complaint	complete exhaustion

Initi

1 Mal

2 App

3 Occ

4 Hei

5 Wei

6 Age

7 Tem

8 Chi

Hist

8.5 Exh

9

10

10.5

11

11.5 St

12

13

14

15

16

17

18

19

POSITIVE FINDINGS CASE 1

Initial Information

1 Male caucasian

2 Appears weak, fatigued, underweight

3 Occupation -- student

4 Height -- 5'10"

5 Weight -- 145

6 Age -- 22

7 Temp -- 99

8 Chief complaint -- complete exhaustion

History of Present Illness

8.5 Exhaustion:

9 Onset -- gradual over a few months, extreme in past two weeks, no precipitating incident.

10 10.5 Character -- generalized weakness and fatigue, no specific weakness

11 Extent -- so exhausted that he needs 10 minute rest after shaving

11.5 Stomach pains and cramps:

12 Onset -- 4 months ago

13 Frequency -- 2-3 times per week initially, now almost every night

14 Duration -- 3-4 hrs, always coming on about 8-9 p.m.

15 Character -- intense, heavy, sharp, "like a big rock"

16 Location (1) -- to right and just below navel

17 (2) -- no radiation, but diffuses over abdomen when very intense

18 Relief -- curling up in a fetal position, aspirin does not relieve pain but relaxes him to permit sleep

19 Exacerbators -- none, no relation to foods

CASE 1

<u>History of</u>

20 Nausea vo

21

22

23 Dizziness

24 Headaches

25 Shortness

26 Chest pai

27

28

28.5 Diarrhea:

29 Onset --

30 Characte

31

32 Frequenc

33 Pain --

33.5 Constipat

34 Onset --

35 Characte

36

Dietary ha

37 Food (1)

38 (2)

39 Fluids -

40 Weight c

CASE 1

History of Present Illness (cont'd)

20 Nausea vomiting (1) -- some nausea at mid a.m. or mid p.m.

21 (2) -- no vomiting

22 (3) -- nausea relieved by food

23 Dizziness -- accompanies nausea mid a.m. or mid p.m. relieved by food

24 Headaches -- occasional frontal headaches

25 Shortness of breath -- on exertion, e.g. walking rapidly, climbing stairs

26 Chest pain (1) -- dull ache in center of chest

27 (2) -- occurs on extended exertion, long walk, climbing 3 flights of stairs

28 (3) -- no radiation of pain

28.5 Diarrhea:

29 Onset -- 3 months ago

30 Character (1) -- loose, not runny stools, brown in color

31 (2) -- mucusy, food particles seen

32 Frequency -- 3 months ago 2 movements per day; recently 4-5 per day

33 Pain -- recently accompanied by gas pain and urgency

33.5 Constipation:

34 Onset -- change to constipation in past week

35 Character (1) -- no change in consistency, just more difficult to pass stool

36 (2) -- mild laxative helps

Dietary habits:

37 Food (1) -- eats regular well balanced meals, good appetite

38 (2) -- needs between meal snacks to allay hunger attacks

39 Fluids -- drinking about 1 gallon per day, mostly milk

40 Weight change -- lost 30 to 35 pounds over 4 month period

CASE

Perso

41 Very

42 Stu

43 Fru

44 Dri

45 Nea

45.5 Mec

Famil

46 Mot

47 Fat

48 Sib

49 Mat

Physi

50 Hea

51 Gen

52

Vit

53

54

55

Abo

56 Sto

57 Sig

CASE 1

Personal Characteristics and Habits

41 Very well organized

42 Stubborn and persistent worries about grades

43 Frustrated by recent disorganization due to illness

44 Drinks beer occasionally, does not smoke

45 Neat, tidy, well dressed

45.5 Medication: aspirin and laxative

Family History

46 Mother -- good health

47 Father -- arthritic about 3 years

48 Sibs -- 2 sisters in excellent health

49 Maternal grandmother dies of diabetes about age 60

Physical Exam

50 Head, eyes, ENT -- conjunctival and mucosal pallor

51 Gen. appearance -- intelligent, polite, well groomed, obviously pale, looks chronically ill, underweight

52 Vital signs: BP -- 120/70

53 Pulse -- 85 & regular

54 Resp -- 19/min

55 Abdomen -- bilateral L.Q. tenderness on moderate palpation, slight guarding on deep palpation. Active bowel sounds with occasional rushes throughout abdomen

56 Stool -- brown, mucusy

57 Sigmoidoscopy -- brown, mucusy stool, pale mucosa

58

59

60

61

62

63

64

65

66

67

68

Laboratory

58 CBC: RBC's 2.56 million (4.6 - 6.2)

Hgn 5.6 gms (14 - 18)

HCT 18% (40 - 54)

WBC 9440 (5000 - 10,000)

Diff Stabs 30%

Segs 55% (54 - 62)

Lymphs 11% (25 - 33)

Monos 4% (3 - 7)

59 MCV -- 70 (80-100)

MCH -- 21.5 (27-33)

MCHC -- 22 (31-37)

60 Reticulocyte 1.8% (.2 - 2.0)

61 Smear Anisolytosis -- moderate

Poikilocytosis -- moderate

Hypochromia -- marked

62 ESR -- 36 mm in 1 hour - corr. for HCT (0 - 10)

63 Prothrombin Time -- 16.5 sec with 13 sec control

64 Stool guiac -- +3 (negative)

65 Stool culture -- mod. growth of hemolytic E. Coli

66 Alk. Phos -- 5.9 BL units (.8 - 2.3)

67 Total Iron -- 32 mcg (65 - 150)

Binding -- 393 mcg (250 - 400)

68 Total Protein -- 8.5 gm (6 - 8)

Alb. -- 4.0 gm (3.8 - 5)

Glob. -- 4.5 gm (2.6 - 3)

69

70

71

72

73

74

75

76

69 Potassium -- 4.8 mEq/l (3.3 - 4.5)

70 Creatinine -- 0.8 mg% (1 -2)

71 U/A L. Amber, cloudy

Sp. Gravity	1.025
WBC	4-5 clumps/HPF
RBC	3-5/HPF
Mucus	+2
Protein	+1
Sugar	Neg
Ketones	Neg
Crystals	+4 amorphous urates
Occult blood	Neg.

Radiology

72 Gall Bladder -- see attached sheets

73 Barium enema -- see attached sheets

74 Upper GI -- see attached sheets

75 Chest -- see attached sheets

76 Abdomen -- see attached sheets

Gallbladder Study

Following the ingestion of a single does of contrast media (9 tablets - 4.5 grams of Telepaque) there is no evidence for concentration of opaque within the gallbladder. The examination was repeated with a double does of dy 18 tablets or nine grams of Telepaque and again no visualization of the gallbladder is demonstrated.

IMPRESSION: Non-visualization of the gallbladder by the double dose technique.

Barium enema

Contrast visualization of the rectal ampulla and sigmoid colon reveals no obvious abnormality in these segments. At the junction of distal and mid thirds of the descending colon slight narrowing of the colon is noted and the margins of the colon become slightly irregular. This change becomes much more apparent in the proximal half of the descending colon and is most obvious throughout the transverse colon where a ragged appearance, indicating mucosal edema and multiple ulcerations, is apparent. The ascending colon is relatively fixed in caliber and the cecal tip appears contracted. The ulcerations are less apparent in the ascending colon and cecum.

The terminal ileum filled readily. Following the post evacuation film, compression spot films of the ileum is demonstrated. The changes described are most consistent with ulcerative colitis (mucosal) involving the right colon and upper 2/3 of the descending colon.

IMPRESSION: Ulcerative colitis involving the entire right colon and proximal 2/3 of the descending colon as well.

Chest: PA and lateral

PA and lateral projections of the chest reveal diaphragm, heart and mediastinum to be normal. There is no pleural abnormality. The lungs are clear and well aerated.

Abdomen

An AP projection reveals minimal scoliosis of the lumbar spine. Gas is noted in the transverse colon. There is no significant bowel distention. No soft tissue masses are noted. No calculus is apparent.

IMPRESSION:
1. Scoliosis involving the lumbar spine.
2. I can't see the psoas shadow as well on the right side. Is there any clinical evidence of peritoneal inflammatory process on the right?

The esophagu

No abnormali

There is sli
is otherwise
extrinsic im
second porti
apparent. I
upon this se
upon the duo
appear to ar
of the dista
posterior in
portion whic
of duodenum

A film made
than 15% gas
the distal j
barium to ai
20% gastric
ileum. At 7
the head of
2 hours the
At 2 1/2 hou
the head of
was refluoro
ileum made.
survey film
is much bet
scoped and
specific ab

IMPRESSION:

UPPER GISERIES & SMALL BOWEL STUDY

CASE 1

The esophagus is normally outlined.

No abnormality of the stomach is demonstrated.

There is slight prominence of the folds within the duodenal bulb which is otherwise normally outlined. A persistent and rather prominent extrinsic impression is demonstrated upon the proximal half of the second portion of duodenum. The significance of this is not readily apparent. It may simply reflect gallbladder impression or impression upon this segment by the right lobe of the liver. It appears to impress upon the duodenum from its posterior and lateral aspect and does not appear to arise from the head of the pancreas. There is slight redundancy of the distal half of the second portion of duodenum which is extremely posterior in position. It may be this unusual position of the second portion which creates the extrinsic pressure upon it. The third portion of duodenum appears normally outlined.

A film made 15 minutes after the ingestion of the barium meal reveals less than 15% gastric retention and the head of the barium column has reached the distal jejunum. The patient was then asked to drink a second cup of barium to aid in the small bowel study. At 45 minutes there is less than 20% gastric retention and the head of the barium column is now in the proximal ileum. At 75 minutes there is only a trace of barium in the stomach and the head of the barium column has advanced slightly in the ileum. At 2 hours the head of the barium column has reached the ileocecal junction. At 2 1/2 hours there is an emptying of the proximal small bowel, again the head of the barium column is at the ileo-cecal junction. The patient was refluoroscoped at this point and compression spot films of the terminal ileum made. The mucosal pattern in the distal ileum appears normal. A final survey film was obtained at 3 hours. At this time the terminal ileum is much better outlined with barium so that the patient was again refluoroscoped and compression spot films of the terminal ileum suggests no specific abnormality. The remainder of the small bowel is normally outlined.

IMPRESSION:
1. No evidence for intrinsic lesion of the esophagus, stomach or duodenum demonstrated.
2. Marked extrinsic impression upon the proximal half of the second portion of duodenum; etiology and significance?
3. No lesion of the small bowel demonstrated.

A fe

comes to

She appea

to have

the foll

occ

hei

wei

age

tem

pul

blo

chi

CASE 2

A female caucasian patient appearing to be in her early 20's comes to your hospital outpatient department at 3 p.m. Monday. She appears pale, fatigued and generally uncomfortable. She claims to have been referred to you be a friend. The nurse has collected the following routine information.

occupation	student
height	5'6"
weight	140
age	23
temperature	102°
pulse	82
blood pressure	120/70
chief complaint	fatigue, poor appetite, headache

Ninete

Female

Present

All sh

only

does

not

slee

has

not

No ap

thr

was

thou

eati
naus

lyir
diff

has

prob
than

thir

been

HISTORY

CASE 2

Nineteen year's old

Female

Present Illness

All she can do is sleep

- only able to stay up 2-3 hours
- doesn't feel really tired
- not refreshed when getting up from sleeping
- sleeps up to 15 hours a day
- hasn't really been sleeping soundly
- not aware of headache when sleeping

No appetite

- three days without eating anything
- wasn't having any solid foods
- thought of food nauseates her
- eating or drinking does not make nausea feel worse
- lying down or sitting makes no difference in nauseated feeling
- has not vomited
- problem is loss of appetite rather than nausea
- thirsty
- been drinking water

CASE

No app

had

woul
to c

was
gett

no c

no a

Heada

thr

usu

loc

do

ve
co

th

do

di
wo

do
t

no

no
in

no

Weak

no

no
on

CASE 2

No appetite (continued)

had a few Cokes

would go specifically to get something
to drink

was getting up and making sure that she was
getting some fluids in her system

no diarrhea

no abdominal pain

Headache

three Excedrin relieves headache for awhile

usual headaches relieved by one aspirin

located right across the front

doesn't spread to the back

very severe -- to the point that she can't
concentrate

throbs

doesn't remember when headache came on

did not go to bed one night feeling well, and
woke up in morning with blinding headache

doesn't notice if they are worse at any particular
time of day

no trouble with vision

not aware of blurring, color, spots
in front of her eyes or dizziness

no change in hearing

Weak all over

no localized weakness

no notice of any particular weakness
on one side of her body or another

CASE 2

Weak all ove

no poor co

Has chills a

has not ha

no particu

Generally ac

maybe a da
of these r

Fair amount

on third d
better --

symptoms w

Getting conc
continue her

little har
three days

In general,
and never ha

Continually
with her usu
good health

Weak all over (continued)

no poor coordination

Has chills and fever

has not had shaking

no particularly heavy sweating

Generally aching

maybe a day or two before the onset
of these really severe symptoms

Fair amount of flu going around

on third day most people are feeling
better -- she feels worse than ever

symptoms worse

Getting concerned whether she could
continue her school work

little hard to focus what it was like
three days ago

In general, been in very good health
and never had any trouble like this before

Continually contrasts how ill she feels now
with her usual general state of vigorous
good health

CASE

CASE 2

Review of Systems

No sore throat

No cough

Does not smoke

Had Laryngitis and cole for couple of weeks (4-6 weeks ago)

Penicillin Pills 4-6 weeks ago for laryngitis

No cold or runny nose now

Little short of breath

But never wakes her up--doesn't wak up panting for breath

Last period was two weeks ago

Periods have been regular

Reasonable number of tampons or sanitary napkins used

Noticed her arms are occasionally broken out around the elbo once the tops of her legs had kind of dry patches.

Had rash when she took some penicillin a few years ago

Earache when a child

Legs were useless--wouldn't function after triple shot of penicillin for earache

Had measles and mumps

Had polio vaccine

Father and younger sister have headaches--usual kind of tension headaches relieved by aspirin. They have them more frequently than the patient did, but don't seem to pay much attention to them.

Cancer may or may not have been involved in mother's miscarriage

Describes herself as bit of tomboy

Is pretty competitive

Once was knocked out while ice skating for a very brief moment

In junior high school a long time ago, and once fell off a bike.

Physical Exam

Pharynx mildly injucted

Tonsils enlarged with small amount of exudate.

color--nondescript, grayish-shite and clear, generally distributed, not in pockets

Conjunctivae reddened

Sclerae slightly reddened and icteric by natural light

CASE 2

Five or six shotty, non-tender nodes palpated in the anterior anterior cervical area to the right of the trachea. Several similar nodes are palpated in both inguinal regions

Breath sounds vesicular at periphery and bronchovesicular centrally with no adventitious sounds

No axillary nodes

Slight tenderness in left upper quadrant of abdomen

Spleen just palpable and slightly tender

Nails show slight pallor

Skin pale

Che

He

CASE 2 LAB

Chemistry

Bilirubin Total	3.45 mg% (.3 - 1.0)
Bilirubin Direct	1.15 mg% (.06 - .25)
Bilirubin Indirect	2.30 mg% (.06 - .80)
B.U.N.	Normal
B.S.P.	10 mg% (0 - 5)
Electrolytes	Calcium, Phosphorus, Sodium, Potassium, Chloride all normal
Glucose Tolerance	Normal
Iron Binding Capacity	390 mg% (250 - 400)
Serum Iron	240 mcg/100 ml (75 - 175)
% Saturation	60% (20 - 50%)

Protein Electrophoresis

Alb	$\alpha 1$	$\alpha 2$	β	Γ
4.2	.30	.47	.64	.95

All normal

Alkaline Phosphatase	2.0 (.8 - 2.3)
L.D.H.	970 (200 - 500)
SGOT	119 units (8 - 40)
SGPT	125 units (8 - 35)
Cholesterol	180 mg% (150 - 250)

Hematology

Hemoglobin	8.9 gm% (12 - 16)
Hematocrit	25% (37 - 47)
W.B.C.	8040/mm3 (5000 - 10,000)
R.B.C.	2.7 million (4.2 - 5.4 million)

Differential

Stabs	Segs	Lymphs	Mono	
13.	24	61	2	%

Autohemolysis

	Control	Patient
Plain	3% (.2 - 4)	35%
With glucose	1.2% (.1 - .6)	3%

Bleeding Time	3 min. 48 sec -normal 1-6
Coag. Time	Normal
Platelet Count	Normal
Prothrombin Time	Patient: 14 sec Control: 13.5 sec
MCV	93 (87 + 5)
MCN	33 (29 ∓ 2)
MCHC	37 (34 ± 2)

Hematolo

Osmoti
Qua

Morph

Retic

Bone
Iro
Leu
Ery
Meg

Microbi

Bloo

Throa

Throa

Serolo

Hete
Slid
Syst
Dire

Urinal

Colo
Char
Reac
Sp.

R.B.
W.B.
Cast

Hematology (cont'd)

Osmotic Fragility, Pres.	Marked hemolysis at .5% NaCl
Quantitative	Increased osmotic fragility pattern
Morphology	Slight basophilic stippling 3 nucleated RBC's seen Atypical lymphs Spherocytes
Reticulocytes	14.3% (.2 - 2.0)
Bone Marrow	
Iron Stain	Positive
Leukocytic Series	Normal
Erythropoitic Series	Hyperplastic
Megakariocytes	Normal

Microbiology

Blood culture	Negative
Throat (gram stain)	Few gram positive cocci in pairs Few gram positive cocci in chains Summary: Normal flora
Throat Culture	+++ Pneumococci ++ Alpha strep ++ Neisseria catarrhalis

Serology

Heterophile, Presumptive	1:224 (<1:224)
Slide Test for Inf. Mono	Positive
Systemic Lupus	Negative
Direct Coombs	Negative

Urinalysis

Color	Amber
Character	Cloudy
Reaction	5
Sp. Gravity	1.013 - normal
R.B.C.	0
W.B.C.	1-3
Casts	0

Urinal

Mucu
Ep.
Bac
Crys

Bile
Urob

Pro
Sug
Occ

Por
Por

Cul

Feces

Occ
Fec

Radic

Che

Ga

Spec

Sp

TB

Urinalysis (cont'd)

Mucus	+
Ep. Cells	+
Bacteria	Negative
Crystals	+ (Amorphous urates)
Bile	Trace
Urobilinogen	++++
Protein	Trace
Sugar	Negative
Occult Blood	Negative
Porphyrins	Negative
Porhpobilinogens	Negative
Culture	No growth

Feces

Occult Blood	Negative
Fecal Urobilinogen	400 Erlich units/100 gm (50 - 300)

Radiology

Chest (P A Film)	Normal
Gall Bladder	Gall Bladder concentration normal. Numerous small stones noted

Special Tests

Spinal Tap	Normal
TB Skin Test	Negative at 48 and 72 hours

himse

He ha

the f

CASE 3

A male caucasian patient with pale complexion presents himself in your hospital outpatient department at 9 a.m. Monday morning. He has had no previous contact with the hospital. The nurse has collected the following routine information on the patient:

occupation	carpenter
height	5'10"
weight	160
age	40
temperature	98.7
chief complaint	left chest pain of 2 days duration

Initial Inform

1 Male Caucas

2 Appears pal

3 Occupation

4 Height -- 5

5 Weight -- 1

6 Age -- 40

7 Temp. -- 9

8 Chief comp

History of P

Chest pai

9 Onset (

10 (

11 Charac

12

13 Locati

14

15 Relie

16 Exace

16.5 Headach

17 Onse

18 Char

19 Chan

20 Loca

POSITIVE FINDINGS CASE 3

Initial Information

1 Male Caucasian

2 Appears pale

3 Occupation -- carpenter

4 Height -- 5'10"

5 Weight -- 160

6 Age -- 40

7 Temp. -- 98.7

8 Chief complaint -- left chest pain 2 days duration

History of Present Illness

Chest pain

9 Onset (1) -- sudden, Friday eve after work

10 (2) -- incurred while wrestling with a friend

11 Character (1) -- sharp, stabbing, intense pain

12 (2) -- continuous since onset

13 Location (1) -- 6th left rib about 4 cm lateral to costochondral junction

14 (2) -- does not radiate

15 Relief -- slight relief from aspirin, sitting still, lying on left side

16 Exacerbators -- deep breath, moving left arm

16.5 Headache

17 Onset -- gradual, about 2-3 months

18 Character -- dull ache, feels like pressure

19 Changes -- always there, says he has gotten to live with it, can ignore it but when he thinks about it, it's there

20 Location -- all over head, no localization

Case **3**

Headach

21 Relie

22 Exac

22.5 Backac

23 Onse

24 Char

25 Loca

26 Reli

27 Exac

27.5 Weakn

28 Ons

29 Cha

30 Cha

31 Weigh

32

33 Medic

Past He

34 Chil

35 Adul

36

37

Habits

38 Alco

39 Smok

40

Headache (cont'd)

21 Relief -- none, tired aspirin but no help

22 Exacerbators -- none

22.5 Backache

23 Onset -- 4 to 5 months ago

24 Character -- occasional, sometimes interferes with sleep

25 Location -- low back

26 Relief -- slight relief from aspirin

27 Exacerbators -- none

27.5 Weakness and Fatigue

28 Onset -- gradual, beginning two months ago

29 Character -- tires easily, generally lethargic

30 Changes -- getting progressively worse, lost several days work during past 3 weeks

31 Weight Loss (1) -- 10 pounds in two months, attributed to lack of appetite

32 (2) -- no specific intolerances

33 Medications -- aspirin only; for chest pain, initially for headache, for backache

Past Health History

34 Childhood -- measles, mumps, chickenpox

35 Adult illnesses (1) -- URI about 4 weeks ago, successfully treated with penicillin

36 (2) -- 6 day course, took it all

37 (3) -- otherwise very healthy

Habits

38 Alcohol -- light to moderate drinker

39 Smoking (1) -- pack a day, for 20 years

40 (2) -- slight morning smokers cough for 2 to 3 years, dry cough

Case

Family

41 Fathe

42 Mothe

43 Sibs

44 Pater

Physica

45 Head

46 Chest

47 Abdom

48 Eyes

49 Throa

50 Back

Lab Tes

51 CBC:

52 Periph

53 ESR --

54 Calci

55 Phos

56 Alk. P

57 L.D.H.

Family History

41 Father -- arthritis for about 5 years

42 Mother -- complaining of vague aches and pains

43 Sibs -- sister and brother in excellent health

44 Paternal Grandmother died of unspecified type of cancer

Physical Exam

45 Head -- several tender spots at various locations on skull

46 Chest -- pain localized to a point on the sixth left rib, tenderness along sixth left rib 3 to 4 cm in extent, occasional premature beats (1-2 per minute)

47 Abdomen -- spleen tip palpable 1 cm. below left costal margin

48 Eyes -- conjunctival pallor

49 Throat -- mucosal pallor

50 Back -- tenderness on lower lumbar spine

Lab Tests

51 CBC: RBC's -- 3.9 million (4.6 - 6.2)
HGN -- 11 gms (14 - 18)
HCT -- 34% (40 - 54)
WBC -- 4100 (5000 - 10,000)
Diff -- Normal

52 Periph smear -- normocytic, normo chromic, rouleau formation

53 ESR -- 35 mm in 1 hour (Wintrobe) (0 - 5)

54 Calcium -- 13.2 mg (9 - 11)

55 Phos -- 2.5 mg (3 - 4.5)

56 Alk. Phos. -- 17 units (5 - 13)

57 L.D.H. -- 425 units (250 - 400)

Case 3

Lab Te

58 Cho

59 SGO

60 Pro

61 Pro

62 Imm

63 Ser

t

a

64 U/A

65 Ben

Radio

66 Sku

67 Che

68 L.S

69. EKG

70 Bon

Case 3

Lab Tests (cont'd)

58 Cholesterol -- 360 mg (150 - 250)

59 SGOT -- 55 units (5 - 40)

60 Prothrombin time -- 15 sec with 13 sec. control

61 Protein electrophoresis -- increased albumin, tall, narrow gamma globulin spike

62 Immunoelectrophoresis -- marked increase in IgM component

63 Serum proteins

total -- 13 gm

albumin -- 5.5 gm

64 U/A -- normal except 4+ protein

65 Bence Jones protein -- present

Radiology

66 Skull film -- multiple punched-out osteolytic lesions

67 Chest film -- pathologic fracture of 6th left rib,
motheaten appearance at fracture,
thinning of bone 5 cm. in extent,
otherwise normal chest

68 L.S. spine -- diffuse mottling in lower lumbar region

69 EKG -- occasional ectopic ventricular contractions (1-2 per minute),
QT internal slightly decreased

70 Bone marrow aspiration -- large numbers of mononuclear cells. These appear to be plasma cells and proplasma cells. Most of these plasma cells display fairly typical nuclear excentricity, perinuclear halo and dark bluish well defined cytoplasm. Nuclei frequently show typical cartwheel pattern but greater variation in size and shape and chromatin pattern than normal. Occasional bi-nucleate or multinucleate plasma cells observed. Many plasma cells appear to be immature. A few mitotic figures.

Case 3

Surgica

of b

Thes

cel

dar

car

pat

cel

fig

71 Surgical rib biopsy -- decalcified sections of bone reveal extensive loss of bony trabeculae with replacement by large numbers of mononuclear cells. These appear to be plasma cells and proplasma cells. Most of these plasma cells display fairly typical nuclear excentricity, perinuclear halo and dark bluish well defined cytoplasm. Nuclei frequently show typical cartwheel pattern but greater variation in size and shape and chromatin pattern than normal. Occasional bi-nucleate or multinucleate plasma cells observed. Many plasma cells appear to be immature. A few mitotic figures.

to yo
accom
They
Her r
be re

c
I
w
a
t
c

CASE 4

A female caucusian patient appearing to be in her mid 20's comes to your hospital outpatient department at 9:30 Monday morning accompanied by her husband. She appears fatigued, pale and overweight. They have been vacationing with another family for the past two weeks. Her regular doctor in another state is also on vacation and cannot be reached. The nurse has collected the following routine information:

occupation	housewife, 1 child age 13 months
height	5'6"
weight	160
age	21
temperature	98.8
chief complaint	nausea and vomiting

Ini

1 F

2 A

3 O

4 A

5 T

6 V

7 O

8 O

His

9 B

9.5 H

10

11

12

13

D

14

15

15.5 N

16

17

18

19

19.5

POSITIVE FINDINGS--CASE 4

Initial Information

1 Female Causasian

2 Appears Pale, Fatigued

3 Overweight (5'6", #160)

4 Age 21

5 Temp. 98.8

6 Vacationing past 2 weeks with another family

7 Occupation -- Housewife

8 One child, age 13 months

History of Present Illness

9 Began feeling "under par" about 5-6 days ago, general weakness and fatigue

9.5 Headache:

10 Onset -- 3 to 4 days ago

11 Character -- sharp, severe, throbing

12 Location -- all over head but manily in back (occipital region)

13 Changes -- worse in a.m., diminshes in intensity during day, progressively worse over past several days

Dizziness:

14 Onset -- accompaning headache

15 Character -- continuous, not severe, vertigo

15.5 Nausea and Vomiting:

16 Onset -- nausea began 2 days ago, got progressively worse, turned to vomiting yesterday, been vomiting with nausea since yesterday.

17 Character (1) -- vomitus described as "Whatever she put in her stomach last"

18 (2) -- no blood, bile or coffee ground material noticed

19 (3) -- non projectile vomiting

19.5 [illegible]

CASE 4

Freque

Intol

Anorexi

Shortne

Onset

Cough (

(

(

Heart R

Medicat

Foods:

<u>Past Heal</u>

Childho

Adult (

(

Surgeri

General

Pregnan

20 Frequency -- every couple hours, 8-9 times since yesterday

21 Intolerance -- no specific food intolerance

22 Anorexia beginning 3-4 days ago

Shortness of Breath:

23 Onset -- 4 days ago, progressively worse

24 PND last two nights

25 now sleeping on 2 pillows

26 Cough (1) -- history of slight, dry smokers cough in a.m., worse in past week

27 (2) -- present cough productive, described as pinkish, frothy, mucusy sputum

28 (3) -- quantity up to 1 tsp. per episode, several times per hour

29 Heart Rate: describes heart as "racing"

30 Medication (1) -- been taking aspirin for headache

31 (2) -- headache worse despite aspirin

32 (3) -- aspirin intake about 8-10 per day for 4 days

33 Foods: Taking only small quantities of mild foods single onset of nausea,

34 trying to take lots of fluids

Past Health History

35 Childhood (1) -- measles, mumps, chicken pox

36 (2) -- no scarlet or rheumatic fever

37 Adult (1) -- flu 2 years ago, not hospitalized

38 (2) -- no serious illnesses

39 Surgeries: T & A at age 8, varicose vein stripped in right leg at age 18

40 General Health: described as excellent

41 Pregnancies (1) -- one full-term pregnancy

42 (2) -- no complications, pre- or post-partum

CASE

43 Al

44 Me

45

46 Co

47 Sm

48

We

49

Pres

50 Va

51 No

52 Ch

2

53 No

54 Re

Fami

55 Mo

56 Fa

57 Si

58 Ma

59 Ot

Physi

60 Hea

d

61 Nec

62 Blo

43 Allergies: Dust and roses (symptoms sneezing, watery eyes)

44 Menstrual (1) -- normal 28 day cycle

45 (2) -- last menses normal, now 3 days late

46 Contraception: Using IUD, inserted 6 weeks post-partum

47 Smoking: 1/2 pack per day, 3 years

48 Weight (1) -- overweight since early teens

49 (2) -- has tried several diets in past, not presently dieting

Present Environment

50 Vacationing in suburban mid-west area

51 No unusual activities

52 Child in family she is vacationing with had sore throat for 3 to 4 days about 2 weeks ago, no other contact with illness

53 No unusual vacation diet

54 Resides in Sandusky, Ohio

Family History

55 Mother -- slight arthritis, age 47

56 Father -- angina, 2 years, age 50

57 Sibs -- two brothers in excellent health

58 Matenial grandmother — died of diabetes about age 70

59 Other grandparents living

Physical

60 Head, Eyes, Ears, Nose, Throat: Unremarkable except pallor and moderate dryness of mucosa and conjunctiva.

61 Neck: venous distension at 45°

62 Blood Pressure: 180/105

P

63

64

65

66

67

68

69

70

71

72

73

L

74

75

76

77

78

79

80

Physical (cont'd)

63 Pulse: 140 and regular

64 Respirations: 28 per minute

65 Chest: dullness at both bases, crepitant rales, decreased breath sounds, holosystolic murmer at 3rd to 4th ICS along left sternal border

66 Abdomen: moderate pain on deep palpation of right abdomen, mild CVA tenderness, rough systolic murmer 2 cm left of umbilicus

67 Pelvic: Normal

68 Rectal: Normal

69 Extremities: Pallor, moderate ankle edema

70 Musculo-skeletal: Normal

71 Nuerologic: Normal

72 Mental Status: Normal

73 Hepato-jugular reflux -- positive

Laboratory

74 CBC:

RBC -- 3.8 million (4.2 - 5.4)

Hemoglobin -- 10 gms (12 - 16)

Crit -- 30% (37 - 47)

WBC -- Normal

Diff -- Normal

75 Morphology -- normocytic, hypochromic

76 Reticulocytes -- .53% ( )

77 BUN -- 37 mg (10 - 20)

78 Creatinine -- 1.8 mg (1 - 1.5)

79 Potassium -- 4.8 mg (3.5 - 4.5)

80 Total Iron -- 50 mcg. (65 - 150)

L

81

82

83

84

85

86

87

88

89

90

91

92

93

F

94

95

96

S

97

98

99

Laboratory

81 Iron Binding -- 410 mcg. (250 - 400)

82 Albumin -- 3.2 gm (3.5 - 5.5)

83 Globulin -- 3.8 gm (2.5 - 3.5)

84 Immunoglobins -- Igm = 150 mg (40 - 120)

85 Cholesterol -- 370 mg (150 - 250)

86 Uric Acid -- 7.5 mg (1.5 - 6.0)

87 Blood Volume -- 47 ml/kg (67, 30% reduction)

88 Electrophoresis --

Alb	α^1	α^2	β	Γ
3.2↓	NL	NL	NL	2.2↑

89 Latex Fixation Titer -- positive

90 C-Reactive Protein -- positive

91 ASO Titer -- 1:320 dilutions, positive

92 Urine

Specific Gravity	1.030 (1.003 - 1.025)
pH	5 (4.6 - 8.0)
Protein	4+ (negative)
24 hr Volume	540 cc ( )
Micro	10-15 RBC casts/HPF, 3-4 hyaline casts/HPF

93 Creat. Clearance 90 cc/min/kg (105)

Radiology

94 Chest x-ray--see report

95 Abdomen--see report

96 IVP-- see report

Special Tests

97 Bone marrow biopsy -- normal except retic. cells 2.1% (.1-2.0)

98 EKG -- sinus tachycardia at 140, tall T wave

99 Thorocentesis -- transudative fluid

CASE 4

Chest Film

Bilateral pleural effusion with marked pulmonary vascular congestion. Normal heart size and contour. No hilar, mediastinal or skeletal abnormalities noted.

ABDOMEN, FLAT PLATE--CASE 2

Psoas shadows well delineated, no fluid level, scattered gases in small intestine. Vertical dimensions of kidneys are: right--16 cm, left--14 cm (upper limit of normal is 13 cm).

No other abnormalities seen.

INTRAVENOUS PYELOGRAM--CASE 2

After the I.V. injection of 50% sodium Hypaque, the dye appeared simultaneously in both kidneys at 1 minute, 3 minutes, 5 minutes and 15 minutes. Vertical dimensions of kidneys are: right--16 cm, left--14 cm (upper limits of normal are 13 cm). Calyces are not distended. At 45 minutes opaque present in the bladder in good concentration. A post-voiding film reveals no significant urinary bladder residual.

APPENDIX C

SUBJECT SCORING FORMS

Please

in di

The m

a. a

b. r

c. c

d. a

Ratin

0 = n

+ = s

t

++ = c

c

- = i

Pleas

neces

APPENDIX C

RATINGS OF DIAGNOSTICITY OF FINDINGS, CASE 1

Please rate each of the numbered findings listed in terms of its helpfulness in diagnosing the major components of the patient's illness.

The major components of the illness are:

a. ulcerative colitis

b. moderate growth of hemolytic E. coli in bowel

c. anemia secondary to malnutrition and blood loss thru the bowel

Ratings

0 = not helpful in diagnosing any of the major components of the illness

+ = somewhat helpful in diagnosing at least one of the major components of the illness

++ = definitely helpful in diagnosing at least one of the major components of the illness

- = inconsistent with at least one of the major components of the illness

Please make the ratings on the numbered rating sheet. If any comments are necessary, please make them on a separate sheet.

Please

in dia

The ma

a. ac

b. re

c. co

d. an

Ratin

0 = ne

+ = se

th

++ = d

o

- = i

Pleas

neces

RATINGS OF DIAGNOSTICITY OF FINDINGS, CASE 4

Please rate each of the numbered findings listed in terms of its helpfulness in diagnosing the major components of the patient's illness.

The major components of the illness are:

a. acute post-strep glomerulonephritis

b. renal hypertension

c. congestive heart failure secondary to hypertension

d. anemia secondary to pulmonary blood loss and reduced erythropoisis

Ratings

0 = not helpful in diagnosing any of the major components of the illness

+ = somewhat helpful in diagnosing at least one of the major components of the illness

++ = definitely helpful in diagnosing at least one of the major components of the illness

- = inconsistent with at least one of the major components of the illness

Please make the ratings on the numbered rating sheet. If any comments are necessary, please make them on a separate sheet.

Please ra
of concer
complicat
the likli
reasonab
metropol

Assume t
no conse

As "anch
(negligi
concern
for risk

Rating of Physical Exam
And Lab Procedures with Respect To
Likelihood and Severity of Risk to Patient Health

Please rate the procedures listed on the 5 point scale with respect to the degree of concern you would have about risk to the patient's health arising from possible complications of the procedures. Your degree of concern should reflect both the liklihood and the severity of risk when the procedures are performed by reasonably competent professionals in a locale similar to the Lansing metropolitan area.

Assume that financial cost and transitory patient discomfort are of absolutely no consequence. (These factors will be evaluated separately.)

As "anchor points" on the scale, consider urinalysis to be rated as 1 (negligible concern for risk); joint fluid aspiration to be rated as 3 (moderate concern for risk); and pneumoencephalogram to be rated as 5 (great concern for risk).

Patient T

Procedure

1. Urinalysis

2. Joint Flui

3. Pneumoence

4. C.B.C.

5. SMA-12

6. Chest Film

7. E.C.G.

8. 24 Hour Ur

9. E.E.G.

0. Lumbar Pun

1. Liver Scan

2. I.V. Pyelop

3. Pulmonary F

4. Arterial Bl

Patient Type: 40 year old male with mild anemia, fatigue and possibility of carcinoma. No obvious cardiac or respiratory problems.

	Physician Concern Over Risk of Procedures				
	Neglig.		Mod.		Great
Procedure	1	2	3	4	5
1. Urinalysis	✓				
2. Joint Fluid Aspiration			✓		
3. Pneumoencephalogram					✓
4. C.B.C.					
5. SMA-12					
6. Chest Film					
7. E.C.G.					
8. 24 Hour Urine					
9. E.E.G.					
10. Lumbar Puncture					
11. Liver Scan					
12. I.V. Pyelogram					
13. Pulmonary Function Test					
14. Arterial Blood Gases					

Procedu

15. Bone M

16. Abdomi

17. Bronch

18. Media

19. Gastr

20. Sigmoi

21. Barium

22. Upper

23. Liver B

24. Surg. R

25. Rectal

Physician Concern Over Risk of Procedures (cont.)

Procedure	Neglig. 1	2	Mod. 3	4	Great 5
15. Bone Marrow Aspiration					
16. Abdominal Exam					
17. Bronchoscopy					
18. Mediastinoscopy					
19. Gastric Tubing					
20. Sigmoidoscopy					
21. Barium Enema					
22. Upper G.I. Series					
23. Liver Biopsy					
24. Surg. Rib Biopsy					
25. Rectal Exam					

Ple
the
enc

Ass
abs

As
(ne
rat
enc

Rating of Physical Exam
And Lab Procedures with Respect To
Patient Pain, Discomfort and Inconvenience

Please rate the procedures listed on the 5 point scale with respect to the degree of pain, discomfort and inconvenience the patient would typically encounter.

Assume that financial cost and possible risk to the patient are of absolutely no consequence. (These factors will be evaluated separately.)

As "anchor points" on the scale, consider urinalysis to be rated as 1 (negligible pain, discomfort or inconvenience); test for blood gases to be rated as 3 (moderate pain, discomfort or inconvenience); and pneumoencephalogram to be rated as 5 (severe pain, discomfort or inconvenience)

Pati

Proc

1. Urin
2. Bloo
3. Pneu
4. C.B.
5. SMA-
6. Chest
7. E.C.G
8. 24 Ho
9. E.E.G.
10. Lumbar
11. Liver
12. I.V. P
13. Pulmona
14. Arteri

Patient Type: 40 year old male, with mild anemia, fatigue and possibility of carcinoma. No obvious cardiac or respiratory problems.

		Pain, Discomfort or Inconvience				
		Neglig.		Mod.		Severe
	Procedure	1	2	3	4	5
1.	Urinalysis	✓				
2.	Blood Gases			✓		
3.	Pneumoencephalogram					✓
4.	C.B.C.					
5.	SMA-12					
6.	Chest Film					
7.	E.C.G.					
8.	24 Hour Urine					
9.	E.E.G.					
10.	Lumbar Puncture					
11.	Liver Scan					
12.	I.V. Pyelogram					
13.	Pulmonary Function Test					
14.	Arterial Blood Gases					

P

P

1. U

2. B

3. Pn

4. C.

5. SM

6. Ch

7. E.

8. 24

9. E.E

10. Lum

11. Liv

12. I.V.

13. Pulm

14. Arte

Patient Type: 40 year old male, with mild anemia, fatigue and possibility of carcinoma. No obvious cardiac or respiratory problems.

		Pain, Discomfort or Inconvience				
		Neglig.		Mod.		Severe
	Procedure	1	2	3	4	5
1.	Urinalysis	✓				
2.	Blood Gases			✓		
3.	Pneumoencephalogram					✓
4.	C.B.C.					
5.	SMA-12					
6.	Chest Film					
7.	E.C.G.					
8.	24 Hour Urine					
9.	E.E.G.					
10.	Lumbar Puncture					
11.	Liver Scan					
12.	I.V. Pyelogram					
13.	Pulmonary Function Test					
14.	Arterial Blood Gases					

Key to List of Physical Examination and Test Procedures

I. Import of Findings Key

1 = noncontributory finding

2 = moderately important finding

3 = critical finding

II. Cost Key

E (Expense): in relative value scale

1 R.V. point = 4 minutes time = $5.00

D (discomfort): in relative value scale

1 R.V. point = Discomfort Equivalent of $5.00

R (Risk): 2 x Relative Value Scale

1 point = Risk Equivalent of $10.00

PHYSICA

Head

Eyes

Ears

Nose

Throat

Appear

Neck

Chest

Abdom

Recta

Pelvi

Sigm

Bloo

Puls

Resp

Extr

Neur

Gen

Gen

Ade

Ski

Bac

* E

	Import of Findings			Cost (All Cases)		
	Case 1	Case 2	Case 3	E*	D	R
PHYSICAL EXAM						
Head	3	2	3	.08		
Eyes	2	2	3	.62		
Ears	1	2	3	.08		
Nose	1	2	3	.05		
Throat	2	2	3	.04		
Appearance, Gross	1	1	3	.04		
Neck	1	3	1	.25		
Chest	3	3	1	1.25		
Abdomen	2	2	2	.50		
Rectal	1	1	2	.21		
Pelvic, Female		1		1.25	5	
Sigmoidoscopy	1	1	2	3.75	7	3
Blood Pressure	1	3	1	.12		
Pulse	1	3	1	.25		
Respirations	1	3	1	.12		
Extremities	1	3	1	.12		
Neurologic	1	1	1	1.25		
Genitalia, Female		1		.04		
Genitalia, Male	1		1	.04		
Adenopathy	1	1	1	.12		
Skin, Hair, Nails	1	1	1	.04		
Back	3	1	1	.62		

* E = Expense in terms of relative value scale where 4 minutes time = 1 R.V. point = $5.00

HEMATOL

CBC (

Smear

Retic

Eryth

Proth

Autol

Blee

Coag

Plat

Osmo

Osmo

Indi

BLOOD

Tota

Albu

Glo

Bil

Bil

B.U

B.S

	IMPORT. FIND			COST (ALL CASES)		
	CASE 1	CASE 2	CASE 3	E	D	R
HEMATOLOGY						
CBC (RBC, WBC, HCT, HGN, DIFF, Indices)	3	3	3	1.3		
Smear for morphology	3	2	3	1.0		
Reticulocytes	2	2	1	1.0		
Eryth. sed. rate	3	2	3	0.6		
Prothrombin time	1	1	2	1.0		
Autohemolysis (plain and with glucose)	1	1	1	1.0		
Bleeding time	1	1	2	1.0		
Coag time	1	1	2	1.0		
Platelet count	2	1	1	1.0		
Osmotic fragility (presumptive)	1	1	1	1.0		
Osmotic fragility (quantitative)	1	1	1	3.0		
Indices	3	2	3	0.0		
BLOOD CHEMISTRY						
Total protein	3	2	2	2.0		
Albumin						
Globulin						
Bilirubin (total)	1	1	1	2.0		
Bilirubin (direct)						
B.U.N.	1	3	1	1.0		
B.S.P. clearance	1		2	0.8		

Bloc

Elec
Ma

Ca

Ph

Sc

Pc

Cl

Bloc

Bloc

Crea

Chol

Iron

Bi

Sa
pH
Enzy

SG

SG

LD

Al

Ac

Ph

Am
Ha
Gl
Ur
Ly
Xy
T_3

	IMPORT. FIND.			COST (ALL CASES)		
Blood Chemistry (cont'd)	CASE 1	CASE 2	CASE 3	E	D	R
Electrolytes						
Magnesium	1	1	1	1.4		
Calcium	3	1	1	1.0		
Phosphorus	2	1	1	1.0		
Sodium	1	1	1	1.0		
Potassium	1	2	1	1.0		
Chlorides	1	1	1	1.0		
Blood sugar random	1	1	1	0.8		
Blood sugar 2 hr post prandial	1	1	1	0.8		
Creatinine	1	3	1	1.0		
Cholesterol	1	2	1	1.0		
Iron Total	3	2	3	3.0		
Binding capacity						
Saturation						
pH	1	1	1	2.0		
Enzymes						
SGOT	1	1	1	1.0		
SGPT	1	1	1	1.0		
LDH	2	1	1	1.0		
Alk. phosphatase	2	1	3	1.5		
Acid phosphatase	1	1	1	1.5		
Phosphokinase	1	1		1.5		
Amylase	1	1	1	1.5		
Haptoglobin	1	1	1	4.5		
Glucose Tolerance	1	1	1	3.5		
Uric Acid	1	2		1.0		
Lypase	1	1	1	1.8		
Xylose Tol.	1	1	1	1.8	3	
T_3 T_4 Thyroid, Each	1	1	1	1.5		

SEROLC

Anti

LE t

Rheu

ASO

R.A.

C-re

Hete

Slid

Dire

VDRL,

URINE

Rout

Urin

24 h

Creat

Pregn

Sterc

Chori

Creat

Tubul

24 hr

Urobil

Porphy

	IMPORT. FIND			COST (ALL CASES)		
	CASE 1	CASE 2	CASE 3	E	D	R
SEROLOGY						
Antinuclear antibody (ANA)	1	1	1	3.4		
LE test	1	1	1	1.4		
Rheumatoid factor	1	1	1	1.0		
ASO titer	1	3	1	1.4		
R.A. (latex fixation)	1	2	1	1.0		
C-reactive protein	1	3	1	0.6		
Heterophile (presumptive)	1	1	1	1.0		
Slide test for mono	1	1	1	1.0		
Direct coombs	1	1	1	1.o		
VDRL, Other for Syphilis	1	1	1	0.8		
URINE						
Routine urinalysis with microscopic	3	3	1	1.4		
Urine for bence-jones protein	3	1	1	0.6		
24 hr volume	1	3	1	0.2		
Creatinine clearance	1	3	1	2.0		
Pregnancy test	1	1	1	1.0		
Steroids; 17 keto, 17 OH	1	1	1	6.0		
Chorionic gonaoatropin (quantitative)	1	1	1	4.6		
Creatine (24 hr)	1	1	1	1.0		
Tubule reabsorption of phosphorus	1		1	4.0		
24 hr urine Protein	3	3	1	0.8		
Urobilogen	1		1	1.0		
Porphyrins, uro-and copro-	1	1	1	4.0		

FECES TE

Parasi

Occult

Urobil

Melani

Fats

RADIOLOGY

Renal

Renal

IV Cho

Skull

Chest

Abdome

IV py

Upper

Bariu

Lumbos

Bone

Gall

Retro

Lung

Femur

Pelvi

Hand
Liver

CULTURE

Sputu

Sputu

Stain

	IMPORT. FIND.			COST (ALL CASES)		
	CASE 1	CASE 2	CASE 3	E	D	R
FECES TESTS						
Parasites or ova	1	1	1	1.0		
Occult blood (guiac)	1	1	3	0.5		
Urobilinogen	1	1	1	2.2		
Melanin	1	1	1	4.0		
Fats	1	1	2			
RADIOLOGY						
Renal Arteriogram	1	1	1	40.0	9	5
Renal Scan	1		1	12.0		3
IV Choleangiogram	1	1	1	8.0	3	3
Skull	3	1	1	3.0		
Chest (AP)	3	3	1	2.0		
Abdomen flat plate	2	2	2	2.0		
IV pyelogram	1	2	1	8.0	2	5
Upper GI and small bowel series	1	1	2	12.0	4	
Barium enema	1	1	3	9.0	5	1
Lumbosacral spine	3	1	1	5.0		
Bone survey	3	1	1	9.0	2	
Gall bladder	1	1	1	8.0	2	
Retrograde IVP	1	1	1	23.8	4	5
Lung Scan	1	3	1	16.0	3	3
Femur	1	1	1	3.0		
Pelvis	1	1	1	3.0		
Hand	1	1	1	2.0		
Liver Scan	1	1	1	12.0		3
CULTURES						
Sputum for cytology	1	3	1	2.0		
Sputum culture	1	1	1	3.0		
Stain, Screening	1	1	1	1.0		

Cult

Pl

Bl

Th

Ur

Ur

Fe

SPECI

Bi

Bi

Bi

Ga

T.

E.

Se
Ur
Th

Ri

Bo

Ly

Sh

Bi

Bi

Bi

Pa

Im

Im

Ga

Lu

Su

	IMPORT FIND.			COST (ALL CASES)		
	CASE 1	CASE 2	CASE 3	E	D	R
Cultures (cont'd)						
Pleural fluid cytology	1	2	1	2.0		
Blood culture	1	1	1	3.0		
Throat culture	1	1	1	3.0		
Urine culture (clean catch)	1	1	1	3.0		
Urine culture (catheter)	1	1	1	3.0		
Fecal culture	1	1	2	3.0		
SPECIAL TESTS						
Biopsy Bowel	1	1	3	23.4	5	5
Biopsy Muscle	1	1	1	20.6	5	
Biopsy-Skin, Mucus Membrane	1	1	1	10.2	5	
Gastroscopy	1	1	1	28.4	3	2
T.B. Skin test	1	1	1	1.0		
E.K.G.	2	3	1	2.4		
Serum protein electrophoresis	3	2	1	3.0		
Urine Protein Electrophoresis	3	3	1	4.8		
Thoracentesis	1	2	1	11.6	6	5
Rib biopsy (surgical)	3	1	1	25.6	9	7
Bone marrow aspiration (sternal)	3	2	3	6.0	7	5
Lymph node biopsy	1	1	1	11.2	9	3
Shillings test	1	1	1	6.0		
Biopsy, liver	1	1		12.6	9	7
Biopsy, lung	1		1	11.6	9	9
Biopsy, kidney	1	3	1	22.8	7	9
Parathyroid assay	1	1	1	8.0		
Immuno Electrophoresis -Serum	3	3	1	4.5		
Immuno Electrophoresis - Urine	3	3	1	5.0		
Gastric Tubing for blood	1	2	1	4.0	7	3
Lumbar Puncture	1	1	1	10.0	6	5
Superficial Buscle Biopsy	1	1	1	20.6	5	0

NAME______________________________

1. Each piece of information requested by the problem solver should be related to a plan of attack for solving the problem. There should be a plan and a well defined purpose behind every question asked.

   a. a commonly heard admonishment to students

   b. not commonly heard but consistent with my training

   c. not commonly heard; faculty would be generally indifferent in opinion

   d. probably controversial; faculty would be divided in opinion

   e. generally inconsistent with my training

2. No diagnostic hypothesis should be more specific or more general than the evidence on hand justifies.

   a. a commonly heard admonishment to students

   b. not commonly heard but consistent with my training

   c. not commonly heard; faculty would be generally indifferent in opinion

   d. probably controversial; faculty would be divided in opinion

   e. generally inconsistent with my training

3. There should always be at least two or three competing hypotheses under consideration at a particular time. Each piece of information should be evaluated with respect to all hypotheses presently under consideration.

   a. a commonly heard admonishment to students

   b. not commonly heard but consistent with my training

   c. not commonly heard; faculty would be generally indifferent in opinion

   d. probably controversial; faculty would be divided in opinion

   e. generally inconsistent with my training

4. Whenever a new or revised hypothesis emerges, the information previously collected (particularly the information from the middle of the sequence of questions asked) should be reviewed. The problem solver should attempt to categorize the previously elicited findings as either tending to confirm or tending to disconfirm his new hypothesis.

   a. a commonly heard admonishment to students

   b. not commonly heard but consistent with my training

   c. not commonly heard; faculty would be generally indifferent in opinion

   d. probably controversial; faculty would be divided in opinion

   e. generally inconsistent with my training

5. When high cost (expensive, uncomfortable or risky) procedures are being considered to confirm a favored hypothesis, the problem solver should consider the possibility of lower cost procedures which might instead rule out one or more diagnostic possibilities in order to make the high cost procedure unnecessary or to increase the probability that the high cost procedure will yield the definitive diagnosis.

   a. a commonly heard admonishment to students

   b. not commonly heard but consistent with my training

   c. not commonly heard; faculty would be generally indifferent in opinion

   d. probably controversial; faculty would be divided in opinion

   e. generally inconsistent with my training

APPENDIX D

SUPPLEMENTAL ANALYSES

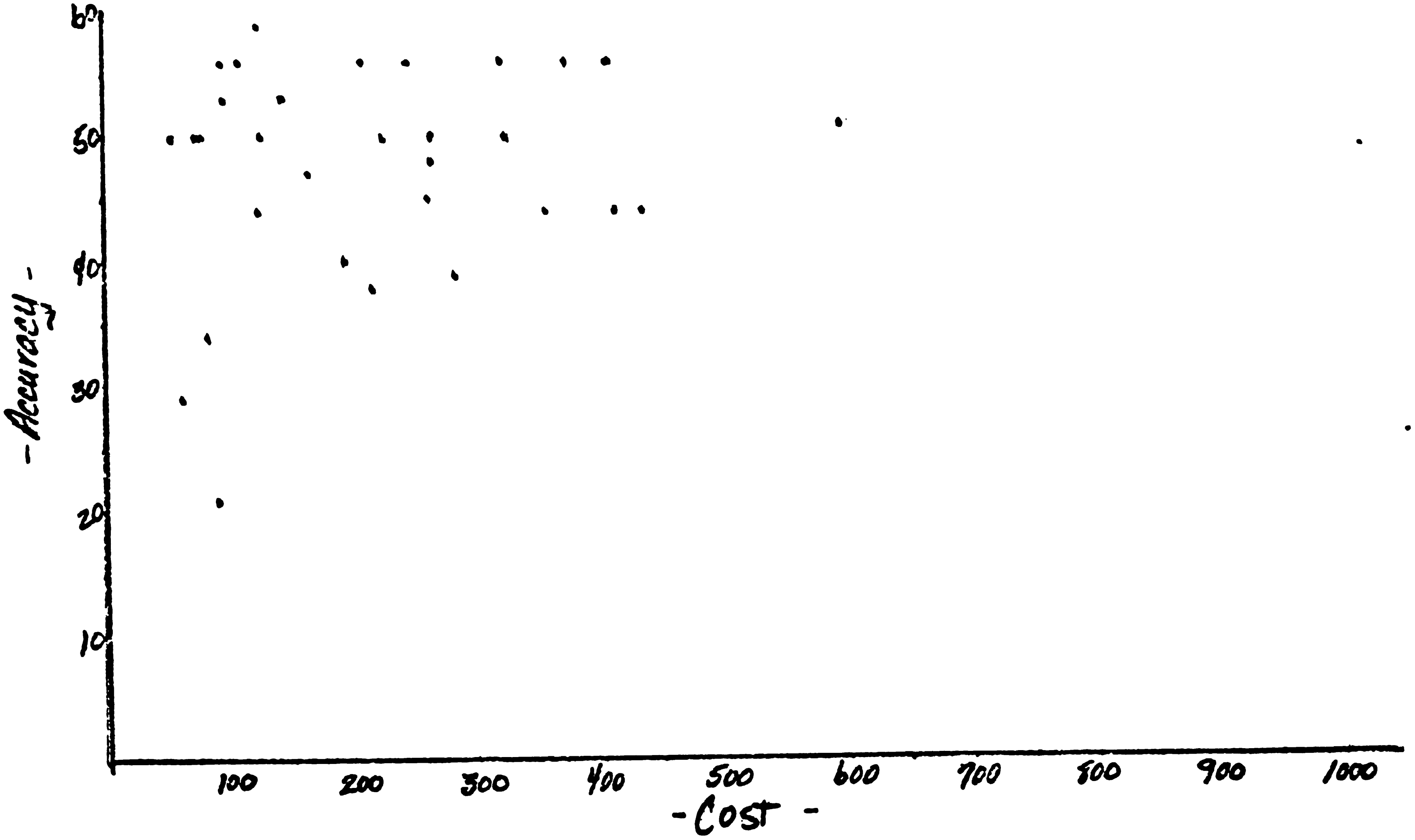

Fig. 7. Scatter Plot of Case 1 Correlation Between Cost and Accuracy

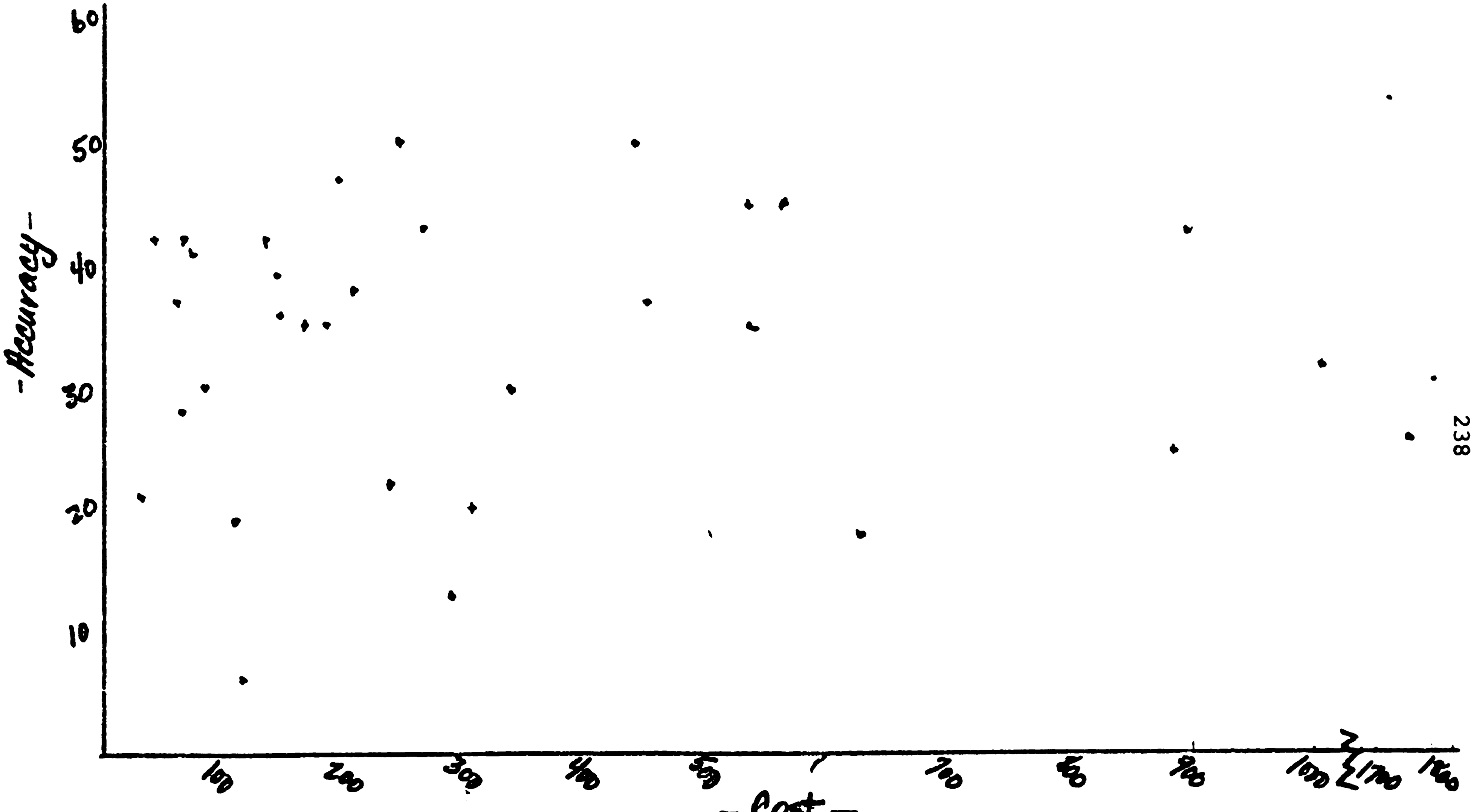

Fig. 8. Scatter Plot of Case 3 Correlations Between Cost and Accuracy

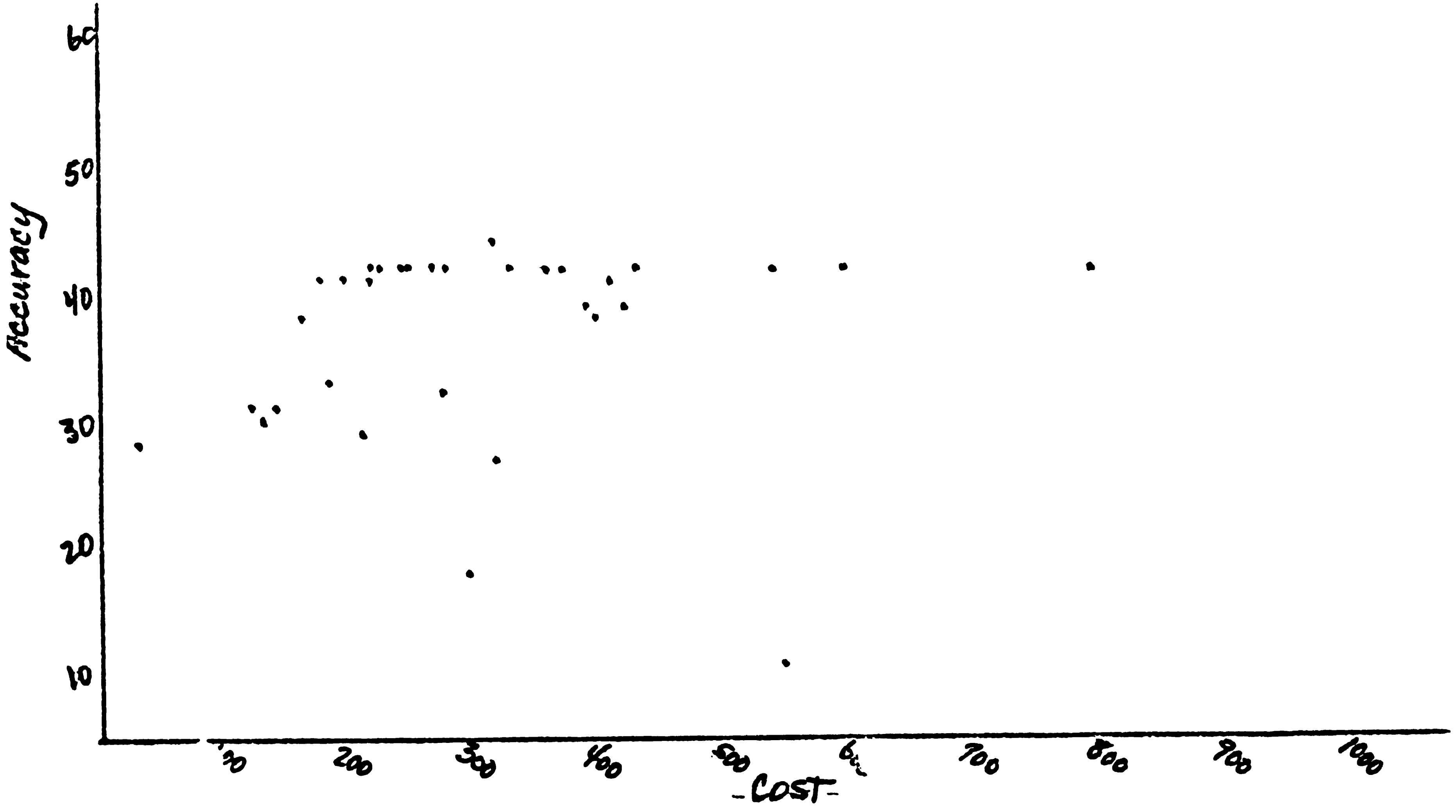

Fig. 9. Scatter Plot of Case 4 Correlation Between Cost and Accuracy

Table 23

Verbatim Idiosyncratic Heuristics of Group 3 Subjects

School 1

Subject 1

a. Good Hx and systematic way of giving it
b. Physical examination
c. Pertinent lab results
d. Tx
e. Patients response to Tx will tell you about disease process
f. Common sense
g. Listen to patient

Subject 2

a. Complete unbiased history and physical
b. Maximize thought process before ordering tests, procedures, ect.
c. Evaluate abnormal values - are they 1^{o} or 2^{o}
d. Plan logical sequence of tests, procedures so that they will not interfere with each other (i.e. IV dye and Thyroid function tests.)
e. Although diagnoses finally seems apparent, do not completely rule out other etiologies.

Subject 3

a. Look at the pt's urine, VDRL, TB test
b. In cases of "exhaustion" remember muscle strength and food fads
c. Anemias - any possible blood loss vs. hemolytic vs. ↓ production
d. Remember family history
e. Remember the skin and hands

Subject 4

a. Keep open mind
b. Organize history - don't jump from system to system.
c. Try to find out how serious the disease is to the pt. Why did he come in now?
d. Don't accept a negative answer without rephrasing it.
e. When you don't know what's happening go to ROS
f. Don't hesitate to get routine V/A & CBC

School 2

Subject 1

a. Listen to what the patient is saying.
b. Make sure you obtain the history in a chronological form, and be firm with yourself in enforcing this rule.
c. Within reason, don't be afraid to re-ask a question to be sure both you and the patient have the facts well in mind.
d. I find it useful to work from "general" differential diagnoses in formulating laboratory tests.
e. If you can try to make your general diff. dx's from the history, in many cases your physical findings & lab tests will serve only to confirm your suspicions.
f. Include pertinent ROS of problems (or dx) suspected within Present Illness as well as in ROS.

Subject 2

a. Get a thorough history & physical
b. Common things occur commonly
c. Occam's Razor
  I. If two events occur in a related time span they tend to be related.
  II. If there are two diagnoses in the previously related events, then the simpler one tends to be correct.
d. If your clinical findings indicate that given lab values do not correlate with them, then your clinical findings are usually the most correct.
e. Listen to the patient, he will tell you what is wrong with him.
f. The sin of commission is worse than the sin of omission or vice versa.

Subject 3

a. Think by systems.
b. Common things are common.
c. Listen to the patient; he'll do the diagnosis.
d. Think by differential diagnosis, formulate diagnosis and proceed to rule out.
e. Pay attention to detail.
f. Don't get lost in detail.
g. Think first--don't spout first thing that comes into head.

Subject 4

a. Complete history
b. Complete physical exam
c. Rectal digital exam
d. Basic studies-- U/A, CBC, Chest film X-ray
e. Confirm hypothesis with relative-friend if patient is not reliable.