ABSTRACT

The underlying idea of this work is that medical students should learn to be effective problem solvers through their curricular experiences. The author studied 32 fourth year medical students from two colleges: 16 were from a college with a traditional curriculum; 16 were from a college whose curriculum included some early attempts at problem-oriented teaching. Students were randomly assigned to one of four treatment conditions. Four students from each college were provided with a set of heuristics for clinical problem solving and prompted to use them; four were provided with heuristics, but received no prompts for their use; four students from each college developed their own personal tests of problem solving heuristics and were prompted to use them; four developed their personal test but received no prompting. Subjects had initially worked through a paper case from which pretest measures were derived. Subjects who received the prepared set of heuristics were given aid in their use through another case. All subjects worked through two additional cases from which posttest measures were derived. The results indicate that fourth year medical students perform better on clinical cases when they have been given powerful heuristics.

(Author/EB)
The ability to solve problems has for some time been proposed as a major goal of schooling. Some of the innovations of the Eight Year Study had their source in the hope that schools could teach students to be more effective problem solvers (Aikin, 1942; Smith, E.R. and Tyler, R.W., 1942). Some of the same hopes and ideas are found in recent curricular innovations in medical schools (Ways, Loftus, and Jones, 1973). The organization of much of the medical school curriculum around particular problems of patients is relatively new. The underlying idea—that medical students should learn an effective problem solving strategy through their curricular experiences—goes back at least to Flexner's famous report (1910).

One of the assumptions that has given problem solving a central place in the curriculum is that problem solving strategies and techniques learned with one problem will be easily transferred to other problems. Thus, problem solving is seen as an activity that has significant generalizable aspects. This view of problem solving has recently been questioned by researchers and curricular designers in medical education. The questions have arisen because studies of clinical problem solving in simulations indicate that performance is case specific (Shulman, 1974; McGuire, 1976). The correlation between performance on one case and performance on another case is low (less than .3, on average). Thus the assumption of transferability of problem solving skills has lacked support.

Continuer findings of low correlations in performance from case to case weaken the argument of those who would make problem solving central in the medical
Such findings suggest that problem solving for a given case depends on knowledge or skill that is more-or-less unique to the condition encountered in the case. If what matters is the possession of specifics relevant to individual cases, the curriculum should emphasize acquisition of as many and as varied a set of specifics as possible. A curricular emphasis on the process of problem solving makes sense only when there is evidence that possession of some tactics or strategies of problem solving leads to improved performance. Re-analysis of a recent study of the role of heuristics in medical problem solving provides such evidence.

The Study

Gordon (1973) studied 32 fourth year medical students from two medical colleges: sixteen students from a college with a "traditional" curriculum; sixteen students from a college whose curriculum included some early attempts at "problem-oriented" teaching. Students in these groups were randomly assigned to one of four treatment conditions. Thus four students from each college were provided with a set of heuristics for clinical problem solving and were prompted to use the heuristics; four were provided with heuristics but received no prompts for their use; four students from each college developed their own personal lists of problem solving heuristics and were prompted to use the heuristics; four developed their personal list of heuristics but received no prompts to use the heuristics. The heuristics provided by Gordon are listed in Figure 1.

Subjects had initially worked through a "paper case" from which pretest measures were derived. Those subjects who received the prepared set of heuristics were aided in using the heuristics as they worked through another case. Then all subjects worked through two additional cases from which post test measures were derived. The four cases, in order, are a problem of exhaustion with a diagnosis of ulcerative colitis; a problem of fatigue and headache with a diagnosis of
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.

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

3. There should 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.

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.

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.

FIGURE 1. The heuristics provided to half of the subjects in the Gordon study.
hereditary spherocytosis triggered by mononucleosis; a problem of left chest
pain with a diagnosis of multiple myeloma; a problem of nausea and vomiting with
a diagnosis of acute glomerulonephritis.

A number of measures are available based upon the record of diagnostic
hypotheses being entertained and actions taken as each subject worked through
each paper case. The variables used in the analyses are: Scope of Early Hypotheses,
Number of Critical Findings Elicited, Cost of History and Physical Examination,
Cost of Laboratory Work, and Accuracy of Definitive Diagnosis. Gordon (1973)
provides a detailed description of these variables and the procedures used in
their measurement.

Original analyses examined four dependent variables, one at a time. Scope,
Critical Findings, Overall Cost (the sum of Costs of History, Physical and
Laboratory), and Accuracy were analyzed in a series of 2 (Schools) by 4 (Treatments).
Analyses of Variance with the pretest scores on each variable used as a covariate.
Only with Accuracy did the analysis suggest statistical significance ($F_{3; 23} = 2.69$;
$p < .07$), but because of the repeated tests a criterion of $p < .01$ had been chosen
as the standard for statistical significance.

A variety of indicators suggested that a reanalysis might be fruitful. First
the costs of laboratory in each case (and consequently the overall costs) are
badly skewed, with mean roughly proportional to standard deviation within cells.
Taking the log of this cost drives the distribution back towards a symmetric
normal distribution and stabilizes the variance. Second, the design of the study
was 2 (School) by 2 (Heuristics) by 2 (Prompting), with real expectation of effects
only for the Heuristics treatment, no expectation of effects associated with
schools, and the effects of prompting uncertain. The dependent variables were
not independent; the direction and strength of their associations could be taken
into account in a multivariate analysis of variance. Finally, with so few degrees
of freedom available it seemed best to use covariates only if they were sign-
ificant predictors of the dependent variables.

As preparation for the analysis the degree of association between the pre-
test variables and posttest variables was tested. It didn't approach statistical
significance. Consequently the posttest variables were analyzed in an analysis
of variance. This is done in two parts: first an analysis of the difference
between scores on the two posttest cases to see whether an interaction of treat-
ment with cases is present; second, an analysis of the sum of scores on the two
posttest cases to determine whether there is an overall effect for the design
factors. The summary statistics for these analyses are displayed in Table 1.

The results in Table 1 suggest a possible sensitivity of treatment effects
to case-specific characteristics. However, interactions are not statistically
significant. The analysis of posttest sums indicates that school and school-
related design factors have no apparent effect on the outcomes of the study. The
results also indicate that experience with and access to the prepared set of
heuristics has an effect on medical problem solving that is likely to be
replicable. The effect can be described by the discriminant function separating
the students who worked with the prepared heuristics from the subjects who worked
with their personal heuristics. The standardized coefficients are indicated in
Table 2. These results are what would be expected from knowledge of the
heuristics. The first, second, fourth, and fifth heuristics urge a more system-
atic and cost-efficient search for and use of diagnostic information. This
should lead to improved accuracy and lower costs and it does. The second and
third heuristic urge the articulation of an appropriately broad set of diagnostic
hypotheses. This should lead to a wider range of early hypotheses, and it does.
The set of heuristics does not bear in an obvious way on the number of critical
findings elicited. The relationship to the critical findings variable is the
weakest in the set.
### TABLE 1

"F" Values Associated With Each of the Design Hypotheses

#### A. For Differences Between Scores on Posttest Cases

<table>
<thead>
<tr>
<th>Design Hypothesis</th>
<th>F Ratio</th>
<th>Probability Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHOOL</td>
<td>1.42</td>
<td>p ≤ .26</td>
</tr>
<tr>
<td>HEURISTICS</td>
<td>2.36</td>
<td>p ≤ .08</td>
</tr>
<tr>
<td>PROMPTING</td>
<td>2.42</td>
<td>p ≤ .07</td>
</tr>
<tr>
<td>HxP</td>
<td>1.06</td>
<td>p ≤ .41</td>
</tr>
<tr>
<td>SxH</td>
<td>.76</td>
<td></td>
</tr>
<tr>
<td>SxP</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>SxHxP</td>
<td>2.29</td>
<td>p ≤ .08</td>
</tr>
</tbody>
</table>

#### B. For Sums of Scores on Posttest Cases

<table>
<thead>
<tr>
<th>Design Hypothesis</th>
<th>F Ratio</th>
<th>Probability Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHOOL</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>HEURISTICS</td>
<td>3.02</td>
<td>p ≤ .03</td>
</tr>
<tr>
<td>PROMPTING</td>
<td>1.27</td>
<td>p ≤ .32</td>
</tr>
<tr>
<td>HxP</td>
<td>1.14</td>
<td>p ≤ .37</td>
</tr>
<tr>
<td>SxH</td>
<td>1.07</td>
<td>p ≤ .41</td>
</tr>
<tr>
<td>SxP</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>SxHxP</td>
<td>1.50</td>
<td>p ≤ .24</td>
</tr>
</tbody>
</table>

*The number of degrees of freedom for each F is 5, 20*
<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope of Early Hypotheses</td>
<td>.72</td>
</tr>
<tr>
<td>Cost of History and Physical</td>
<td>-.45</td>
</tr>
<tr>
<td>Cost of Laboratory Work</td>
<td>-.46</td>
</tr>
<tr>
<td>Number of Critical Findings Elicited</td>
<td>-.29</td>
</tr>
<tr>
<td>Accuracy of Definitive Diagnosis</td>
<td>.96</td>
</tr>
</tbody>
</table>
Implications

Clearly the critical experimental effect—the effect of the set of prepared heuristics—is not so large as to be inescapably obvious in analysis. In order to find the effect you have to look for it with care. The effect found, however, is precisely the effect expected and that should redouble our confidence that it will be robust and stable.

The results indicate that fourth-year medical students perform better on clinical cases (more accurately and economically) when they have been given powerful heuristics. This effect appears despite the relatively brief introduction (experience in one case) that the experimental subjects had to the heuristics. It appears despite the fact that since the subjects were in their fourth year of medical school they were not clinical novices but had had time to develop adaptive strategies of medical problem solving. And it appears despite the background of "noise" that case-specific effects presents to the analysis.

There can be effects of strategy in clinical problem solving. The effects are seen most clearly when students are provided with more efficient problem solving strategies. Effective problem solving heuristics appear to be teachable. The implication seems inescapable: medical curricula should provide opportunities to learn powerful strategies for problem solving.

The Gordon study is suggestive in other ways. Correlations between performance on one case and those on another case were low in this study as in others (see Table 3). In correlations alone we find no grounds for support of attempts to teach problem solving in the curriculum. The Gordon study suggests that we will learn the value of intervening in the curriculum to teach problem solving only by carefully controlled study of such interventions. Curriculum studies will have to become more experimental and interventionist, less passive and correlational. Correlational study can complement but cannot supplant the needed experimental studies.
TABLE 3

Averages of Between-Case Correlations* Among Problem Solving Variables.

<table>
<thead>
<tr>
<th></th>
<th>Scope</th>
<th>Critical Findings</th>
<th>History &amp; Physical Costs</th>
<th>Lab Costs (Log Transform)</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Critical Finding</td>
<td>.09</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History &amp; Physical</td>
<td>.23</td>
<td>.38</td>
<td>.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lab Costs</td>
<td>-.11</td>
<td>.25</td>
<td>.27</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>-.10</td>
<td>.10</td>
<td>.10</td>
<td>.05</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Correlations based on deviations from cell means
The Gordon study also suggests two strategies that should prove useful in moving us closer to an understanding of the process of problem solving in important practical situations. A first strategy might be called disaggregation—the development of a comprehensive set of narrowly defined variables that can be aggregated in various ways. This provides the analyst freedom to look for forms of aggregation that provide the most stable characterization of problem solving behavior. For example, Gordon monitored seven different elements that contribute to the cost of a work-up—the direct expense of a history, a physical examination, laboratory tests; the risks accompanying a physical examination, laboratory tests; and the discomfort associated with a physical examination, laboratory tests. These seven can be aggregated into the costs of a history and physical and of laboratory tests as in the present analysis. Or they can be aggregated in other ways that may reveal hidden but stable aspects of clinical problem solving performance.

The second strategy is an attempt to capture the constraints on the problem solver and his intentions as accurately and comprehensively as possible. The Gordon study makes major steps towards a fuller recognition that the clinical situation presents choices of action that involve both costs and a probability of benefit. Currently, in both the medical curriculum and studies of clinical problem solving the practitioner's problem is presented as one of reaching an appropriate conclusion—i.e., a correct differential diagnosis. This substitution of a problem of knowing for a problem of acting is characteristic not only of the medical curriculum, but of curriculum generally. The instrumentation of the Gordon study suggests what may be required if we are to more accurately represent the problem faced by the clinician (or of the man of practice, in general). It is essential that we capture the full range of practical elements in the "space" of a problem if we are to develop more adequate ideas about what constitutes functional problem solving ability.
REFERENCES

Aiken, W.M. *The Story of the Eight Year Study*, New York: Harper. 1942

Flexner, A. *Medical Education in the United States and Canada*, Bulletin No. 4, Carnegie Foundation for the Advancement of Education, New York. 1910


