FACTORS IN THE EARLY DEVELOPMENT AND EVALUATION OF CLINICAL COMPETENCE IN MEDICAL STUDENTS

Hugh M. Scott, M.D., Martin Wong, Ph.D.,

and

Peter Ways, M.D.
Factors in the Early Development and Evaluation of Clinical Competence in Medical Students

INTRODUCTION

"In the end the final test of a Medical School is its outcome in the matter of clinicians."

Abraham Flexner, 1910

This "ultimate criterion" of Medical School success by Flexner has remained valid over the intervening 60 years. However, in more recent times dissension has arisen as to the applicability of the means he proposed to achieve this end in terms of curriculum. This has led to a proliferation of "new curricula", particularly in the past decade. The evaluation of these new curricula has rarely been done systematically and therefore it is difficult to judge their degree of success. The design of any such evaluation must consider the root causes of this ferment in Medical Education, to which the curricula are responding. A few of the most frequently cited include:

A.) The rapid and massive expansion of knowledge relevant to Clinical Medicine has forced a grudging acceptance by Faculties that it is unrealistic to expect students to "learn it all".

B.) What is considered "knowledge" is changing so rapidly in some areas that it is difficult to predict what students should be encouraged to learn. It is a corollary that they must learn in such a way that a lifetime of relearning and new learning will enhanced.

C.) The multiplicity of career roles in Medical School graduates make it unlikely that all students will require the same knowledge.

D.) In the interest of efficiency, knowledge which students do acquire should be assimilated by them in a manner which is most likely to be useful in the solution of clinical problems. A strict disciplinary orientation may not be the best approach to achieve this objective.

These dilemmas and problems are not, of course, unique to Medical Education. Educational Psychologists in general have begun to recommend a greater emphasis on the learning of intellectual processes and strategies rather than "facts" as an at least partial solution to them. This attitude has received increased impetus and crystallization through Bloom et al.'s "Taxonomy of Education Objectives-Cognitive Domain". Similarly, alternatives in instructional strategy, exemplified by "Discovery Learning", have been suggested to better achieve such "Higher Process" learning.

HUGH M. SCOTT, M.D. MARTIN WONG, PH.D. AND PETER WAYS, M.D. College of Human Medicine, Michigan State University.
These were among the considerations leading to the development of the curriculum for the transitional year (combining basic science and clinical emphasis) in the Michigan State University College of Human Medicine. The programme is divided into 2 parts. The initial two thirds consists of 18 "focal problem exercises". Each focal problem is a common clinical paradigm (e.g. chest pain). For three days, faculty-student groups explore the mechanisms and effects of this paradigm exemplified by simulated and actual cases.

This is usually the means of presentation of "problems" to the physician by patients. It appears to be a form of learning setting which better approximates the application setting of Clinical Problems Solving than an organ system, disease or discipline orientation.

Every fourth week is devoted to evaluation and review. The final third of the year consists of a general clerkship, in which the same instructional orientation is encouraged.

Several crucial evaluation questions arise in connection with the institution of such a curricular format.

A.) Entering Characteristics

Most admission committees rely to a greater or lesser degree on college grade point average (G.P.A.) and Medical College Admission test (M.C.A.T.) scores. The former is not derived from nor was the latter designed for prediction of performance during this type of curricular experience. It is unclear how much either index reflects intellectual process development as opposed to content acquisition. Therefore, are these still the most appropriate instruments for predicting performance or must a new means for prediction be found? Would some combination of Cognitive Psychological tests more specifically tend to reflect problem solving aptitude be superior?

B.) Performance during the Focal Problem Sequence

Bloom has pointed out that, "achievement examinations effect the methods of study students use and they appear to have significant consequences for the learning process.... For example, suppose the learning exercises are primarily concerned with complex types of problem solving and examinations stress only recall of specific information? Does...he (the student)develop only the characteristics required by the examinations".
Also, in their studies of the evaluation of clinical proficiency, both Helfer and Slater\(^9\) and McGuire\(^6\) have noted the low correlation between measures of content and measures of process despite relatively high correlations between independent measures of process.

It therefore appears essential that measures of process be central in any evaluation of a curriculum with such an emphasis. However, as content is learned in association with process in this curricular format, does the previously reported dichotomy between measures of process and content persist?

C.) Performance during the Clerkship

Although these were still junior students, the clerkship represented their first extended period as Clinicians. Due to the degree that early learning is the most important learning in any field, their performance during the clerkship represented the beginning of Flexner's "ultimate criterion". The objective of the entire curriculum was to solve Clinical Problems. Which, if any, of the previous measures predicted this ability during the clerkship?

This paper describes the results of preliminary efforts to answer these questions with the first class of 25 students experiencing this curriculum.

MATERIAL AND METHODS

A.) Entering Characteristics

M.C.A.T. and G.P.A. scores were available on all students and were converted to standard scores for correlation purposes. (The sum of the 4 M.C.A.T. sub-tests was used). The model of problem solving described by Dewey was used to select 3 cognitive psychological tests to derive a "Problem Solving Index". (His model suggests that problem solving involves at least 4 distinct processes; Problem sensing, problem formulation; search behaviour and problem resolution). As an indicator of problem sensing; 3 sub-tests were administered (picture completion, picture arrangement and seeing deficiencies), and the scores combined for a composite total; As an indicator of problem formulation, the Watson-Glaser test of Critical Thinking was used; as an indicator of potential for creative search behaviour, Mednick's Remote Associates test was used.

Standard scores were derived for each category and the 3 summed to provide the "Problem Solving Index", (P.S.I.)
B.) **Performance during the Focal Problem**

During each of the 6 evaluation weeks measures of both content and problem solving acquisition were taken. The content measures were predominately objective and short answer examinations. The measures of process were designed so that during each evaluation week measures of problem sensing, problem formulation, search and problem resolution (diagnosis and/or decision) were taken. Standard scores were derived for each sub-test and summed across the first and second 3 evaluation weeks respectively. This provided a composite score for both process and content acquisition during the 2 halves, as well as a grand total for both.

C.) **Performance during the Clerkship**

The evaluation of Clinical Performance remains somewhat of an enigma. Optimaly, a measure of habitual or chronic behaviour is desirable. Typically this has been provided by supervisor's ratings. However, the sources of potential invalidity in such ratings are well known. Patient records written by the students appear to offer an alternative form of evaluation. Weed has suggested a modification of patient records into a "Problem Oriented Format". This provides a demonstration of the Clerk's analysis of the patient's clinical problem in a manner which closely parallels Dewey's formulation of general problem solving. Students were instructed and directed to keep records in this format. Upon completion of the clerkship, 4 charts for each student were selected in a stratified random fashion (2 from the first half and 2 from the second half), from all the records produced during the clerkship. Each record was then analysed by 2 of a team of 4 clinical faculty according to a standardized scale of problem analysis and resolution. (Each faculty member reviewed a record from each half of the clerkship for each student and no pair of faculty examined more than one chart together of any given student). The same 4 faculty evaluated all records. A mean score for each record was derived and the 4 means totaled to provide an index of clinical problem solving for each student.

Finally, for purposes of comparison with an "acute" evaluation method of performance, each student completed a "Diagnostic Problem Exercise" of the type first described by Rimoldi and modified by Helfer and Slater. The "Competence Score" as described by Helfer was derived for each student. All scores were converted to standard scores for purposes of Correlation.

**RESULTS**

The Correlations relevant to the questions raised in the "Introduction" are outlined in Table 1. (Only those considered relevant to this discussion are shown, rather than the complete matrix).
DISCUSSION

A.) Entering Characteristics

Neither M.C.A.T. nor P.S.I. provided any meaningful prediction of performance, either during the focal problem exercise nor during the clerkship. Indeed, some of the M.C.A.T. correlations are strikingly negative (-0.35 with both total process and the diagnostic problem exercise). Without wishing to prolong the longstanding debate concerning the value of the M.C.A.T., it does appear that it requires modification to be useful for admission to a programme such as is described here.

It had been hoped that the P.S.I. would offer an alternative or at least a partial one. This did not happen. Clearly prediction of adjustment to such a self-learning oriented curriculum as this one requires other characteristics not measured by these tests.

The G.P.A. was the most successful of the 3 although only in the prediction of total content score ($C_t$) did it approach meaningful significance ($0.41$).

B.) Focal Problem Performance

Contrary to the experience noted above, significant correlations between measures of process and content were found. (Total Content vs. Total Process = 0.70, $P_1$ vs $C_1$ = 0.64, $P_2$ vs. $C_2$ = 0.31). It is interesting to note as well that the process scores showed significant stability between the 2 halves ($P_1$ vs. $P_2$ = 0.40), but the content scores did not ($C_1$ vs. $C_2$ = 0.00). These sorts of findings are easily over interpreted. However, it is tempting to at least hypothesize that under this curricular format where content is learned with a process orientation, their acquisition is closely parallel. The previously reported low correlation were from different curricular formats.

The instability of the content measures may represent differential student ability to learn on their own with the stimulation provided by the problem solving curricular orientation, whereas the process learning took place largely in the classroom setting and was thus more stable. This stability suggests that there may be some underlying personality and/or intellectual characteristics which provide aptitude for learning in this setting.

C.) Clerkship Performance

The low correlation between Record Analysis scores and scores on the Diagnostic Problem Exercise (0.16) may at first suggest unreliability of one or both measures. However, another interpretation is possible. Namely, the D.P.E.'s represent what the
students could do, the Record analysis what they did do. In any event, both total process and total content scores did give a good correlation with the D.P.E. (both 0.63). Correlations with record scores are not in the range of significance. One can only suspect that this lies outside the range of prediction of intellectual scores to include affective variables, and also the differential response by individual students to impending National Board Examinations.

CONCLUSION

From these preliminary findings, it does appear that when repeated measures are used, the dichotomy between process and content acquisition is not great using this curricular format. Both appear to provide meaningful prediction for aptitude for clinical medicine. Measurement of performance in the clinical setting is still a problem but we feel that Record Analysis deserves further study.

However, the study clearly provides no solution to the problem of student selection. Indeed both the institution of such new curricula and also the national desire to enroll students from educationally deprived backgrounds (leading often to poor M.C.A.T. performance) make it imperative that more meaningful predictors of performance than those currently available be sought.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Dr. Ray Helfer of Catholic Hospitals, New York and Mr. Dennis Hepting of the University of Colorado for the computer scoring of the Diagnostic Problem Exercises. Also to Dr. R.W. Waddell of McGill University for statistical analysis of the data.
BIBLIOGRAPHY


8.) Weed, L.L., Medical Records that Guide and Teach, New England Journal of Medicine, 1968, 278, 593-600, 652-657.

<table>
<thead>
<tr>
<th></th>
<th>PSL</th>
<th>C₁</th>
<th>P₂</th>
<th>Pₜ</th>
<th>Cₜ</th>
<th>DPE</th>
<th>Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>-0.21</td>
<td>0.12</td>
<td>0.41</td>
<td>0.20</td>
<td>0.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCAT</td>
<td>0.08</td>
<td>0.35</td>
<td>-0.10</td>
<td>-0.35</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P₁</td>
<td>0.64</td>
<td>0.49</td>
<td>0.70</td>
<td>0.63</td>
<td>0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C₂</td>
<td>0.00</td>
<td>0.31</td>
<td>0.09</td>
<td>0.63</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.P.E.</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rec.</td>
<td>-0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P₁ = Process Measures, 1st 3 weeks
P₂ = Process Measures, 2nd 3 weeks
Pₜ = Combined total, all process measures
C₁ = Content Measure, 1st 3 weeks
C₂ = Content Measure, 2nd 3 weeks
Cₜ = Combined total, all content measures.

Rec = Mean Record Analysis Score