DOCUMENT RESUME

ED 195 178 HE 013 157

AUTHOR Carey, Christine Regan, Ed.
TITLE Research in Medical Education: 1980. Proceedings of the Nineteenth Annual Conference.
INSTITUTION Association of American Medical Colleges, Washington, D. C.
PUB DATE 80
NOTE 407p.
AVAILABLE FROM Association of American Medical Colleges, Suite 200, One Dupont Circle, Washington, DC 20036

EDRS PRICE MF01/PC17 Plus Postage.
DESCRIPTORS Academic Achievement: Academic Standards: Agency Role: Career Choice: *Clinical Experience: College Environment: *Graduate Medical Education: Higher Education: *Medical Education: Physician Patient Relationship: Professional Continuing Education: Rural Areas: Teacher Effectiveness

ABSTRACT

The proceedings categorize 55 papers on research in medical education in 12 sections: standards and standard setting (including equivalence-testing methods, standardized examinations, and foreign exams): predicting career outcomes (career choice and satisfaction): teaching clinical skills (student-patient relations and interview skills): teaching in the clinical setting: continuing medical education and participation in it: measurement perspectives on the impact of residency programs (faculty perceptions, test performance, and patient instructors): enhancing student/faculty environment interactions (medical student needs, affective learning, study skills counseling and techniques, and a program to improve instructional skills): evaluating continuing medical education: clinical problem solving instruction: the predictability of predictive techniques of student performance: development of patient attitudes (in medical students, primary care physicians, and intensive care unit residents): and symposia. Topics in the last category include definitions of competence in graduate medical education: cost awareness education: educational models in primary care: education for rural practice: clinical reasoning: teaching patient examination skills by simulation: rural and primary care physician recruitment and selection: medical school learning environments: senior clerkships: and notes on the standards developed by the Joint Committee on Standards for Educational Evaluation. The conference schedule and list of participants are appended. (MSE)

RESEARCH IN MEDICAL EDUCATION: 1980

PROCEEDINGS

OF THE

NINETEENTH ANNUAL CONFERENCE

Sponsored by
ASSOCIATION OF AMERICAN MEDICAL COLLEGES
Division of Educational Measurement and Research
in conjunction with the 91st Annual Meeting
October, 1980

NINETEENTH ANNUAL CONFERENCE

ON

RESEARCH IN MEDICAL EDUCATION

Washington, D.C.

Washington Hilton Hotel

October 29, 1980

PLANNING COMMITTEE

FOR THE

NINETEENTH ANNUAL CONFERENCE ON RESEARCH IN MEDICAL EDUCATION

CHAIRMAN: Gary M. Arsham, M.D., Ph.D.	Pacific Medical Center
Richard M. Caplan, M.D.	University of Iowa
Joseph S. Gonnella, M.D.	Jefferson Medical College
Kaaren I. Hoffman, Ph.D.	University of Southern California
Frank Schimpfhauser, Ph.D.	State University of New York-Buffalo
Frank T. Stritter, Ph.D.	University of North Carolina
James B. Erdmann, Ph.D., Secretary	Association of American Medical Colleges

Proceedings Compiled By: Christine Regan Carey

FOREWORD

As the Conference on Research In Medical Education approaches its 20th anniversary, many of you may have noted some recent changes in the review process for papers and symposia, in the conduct of the conference, and in the form of this document itself.

- All papers are reviewed in a masked fashion to minimize reviewer bias.
- Reviewers' comments are provided to all accepted authors, paper session moderators, and symposium organizers, as well as to those "rejected" authors and symposium organizers who request them.
- The Proceedings are indexed in the Index Medicus.
- The Proceedings are copyrighted, primarily to protect the authors (permission will routinely be granted to those authors who wish to use their presentations in another setting or format, and the AAMC will always request author approval if a third party should request the use of their paper).
- The role of the paper session moderator as a discussion leader has been strengthened to provide analytic comments about the papers presented, raise critical questions and to offer several minutes of discussion and summary to provide a stimulus for further questions and discussion by the audience.

With regard to the review and selection process, many of you have asked how the committee operates. The committee makes no effort to guide the topic or subject matter of the submissions. However, the committee has been increasingly more rigorous in its application of the selection criteria that are outlined in the call for papers. The Committee judges the importance and significance of the project, the quality of the research design, and the usefulness of the results to others. While there were many close decisions, problems often found among those papers that were rejected included inadequate research designs, lack of clear hypotheses, minimal or very preliminary data, or little generalizability or utility to others. Many of the last were descriptive internal evaluations of particular programs. The committee clearly recognizes that institutional resources are increasingly directed toward institutional decision making, rather than generalizable knowledge. Such projects, however, should be presented in the light of how they are of interest and importance to other institutions or programs. On the other hand, certain papers have been selected not because they were without flaws but because the committee felt that the basic project or research was worthwhile and could profit by an open and constructive discussion among peers.

With respect to the range of topics submitted this year, the committee noted a larger number of papers on continuing medical education, as well as on evaluation of primary care programs. Certainly, this reflects at least in part the increased institutional resources that are being directed at these types of programs. The committee also noted that there were more studies of methodology and methodological issues than in recent years. There were fewer innovative projects described, probably because of the new forum for demonstrations and innovations in medical education offered through the Group on Medical Education. However, it is certainly appropriate for evaluative studies of innovations, curricular changes, or instructional development projects to continue to be submitted to RIME. Unfortunately, with few exceptions, no clear paths of research are being followed in an effort to lend generalizable new knowledge to the field of medical education. Clinical problem solving and decision making represent the clearest example in which earlier research is being built upon and guided by prior work.

The committee views this conference as much more than an opportunity to hear presentations. If that were the sole goal, no conference would be necessary and these Proceedings would suffice. The annual meeting is intended as a learning opportunity for presenters as well as the audience, an opportunity for scholarly interaction to test ideas among peers.

The committee is eager to continue to improve the annual conference and to make it responsive to your needs. A member of the committee observes each session of the conference. We certainly welcome your written feedback as well, not only on the conference, but also on the review and selection process, and this document.

Gary M. Arsham, M.D., Ph.D.
For the Committee

TABLE OF CONTENTS

PAPER PRESENTATIONS

STANDARDS AND STANDARD SETTING

PREDICTING CAREER OUTCOMES

TEACHING CLINICAL SKILLS

TEACHING IN THE CLINICAL SETTING

PLANNING CONTINUING MEDICAL EDUCATION

IMPACT OF RESIDENCY PROGRAMS: MEASUREMENT PERSPECTIVES

ENHANCING STUDENT/FACULTY ENVIRONMENT INTERACTIONS

EVALUATING CONTINUING MEDICAL EDUCATION

CLINICAL PROBLEM SOLVING

PREDICTABILITY OF PREDICTIVE TECHNIQUES

11

DEVELOPMENT OF PATIENT ATTITUDES

SYMPOSIA

13

PRECIS

STANDARDS AND STANDARD SETTING

MODERATOR: Bryce Templeton, M.D.
National Board of Medical Examiners

THE USE OF THE RASCH MODEL TO TEST THE EQUIVALENCE OF TWO METHODS OF STANDARD SETTING

The Rasch latent trait model was used to analyze student responses to a criterion-referenced test which consisted of two item types and a minimum performance level determined by one of two item types. The results indicated that the method used to set a criterion score is significant in determining the educational outcome.

A COMPARISON OF SERVERAL SCORE CUTTING PROCEDURES AND THEIR EFFECTS ON SUCCESS RATES

In this study, normative and absolute standard procedures for setting a cutting score on a medical certifying examination are compared. Results indicate that the cutting scores produced by the various methods are similar, but the variation in the number of candidates failing, according to procedures, is more pronounced.

CONTENT REPRESENTATIVENESS AND STUDENT PERFORMANCE ON NATIONAL BOARD PART I SPECIAL SUBJECT EXAMINATIONS

The content representativeness of selected NBME Part I special subject examinations (subtests) is assessed and correlations between faculty ratings of the teaching emphasis given to the content of NBME test items and student performance on those items are presented. The results are discussed in terms of the usefulness of external medical examinations to evaluate intramural student learning.

THE FEASIBILITY OF USING A CANADIAN EXAMINATION FOR THE CERTIFICATION OF AUSTRALASIAN CANDIDATES

A multiple choice examination used for certifying candidates eligible for certification in the specialty of Psychiatry was administered to a group of Canadian candidates seeking certification in the same specialty. Results from the test administrations indicated the performance of Canadian and Australasian candidates was very similar.

THE USE OF THE RASCH MODEL TO TEST THE EQUIVALENCE OF TWO METHODS OF STANDARD SETTING

Harasym, P. H.[1], Faculty of Medicine, University of Calgary

Criterion-referenced testing relates student performance to absolute standards rather than to the performance of other examinees. Educational decisions (i.e. student progress, promotion, certification, and/or graduation) are based on whether a student's performance is above or below an absolute standard. Central to this practise is the process of standard setting. If the standard is set too high or too low, inappropriate and perhaps damaging decisions can be made. The major characteristic of the standard-setting process is that it is fundamentally judgemental and results in arbitrary standards. This aspect of the process is heavily criticized. Glass (1978), in his analysis of the merits of standard-setting, concluded that

> *Setting performance standards on tests and exercises by known methods is a waste of time or worse* (p. 259).

In her analysis, Burton (1978) concluded that the practise is without a practical technology. In spite of its criticisms, the usefulness of the practise is widely upheld (Millman, 1973; Messick, 1975, Linn, 1979, Jaeger, 1979, Sheppard, 1979, and Hambleton, 1980) and the movement has spread throughout North America.

Several techniques are available for setting standards. Glass (1978) identified six classes of techniques for standard setting. Different types of multiple-choice items are also currently being used in testing. Hubbard (1978) reports that the National Board of Medical Examiners uses three basic item types: one-best-response type (item type A), the matching type (item types B and C), and the multiple true-false type (item type K and X). With the many item types and standard setting procedures, little is known as to the extent to which a given item type and procedure will yield consistent results. Because of the lack of knowledge in this area, practitioners are often at a loss as to which item type and procedure to select. Hambleton (1980) points to the need for greater research in this area and calls for more empirical investigations to determine which factors should be considered in arriving at a decision about the standard-setting method to be used in particular testing situations. He also states that the many standard-setting methods need to be described accurately, and their advantages and disadvantages noted.

The objective of this investigation was to determine empirically whether two multiple-choice item types and two corresponding standard-setting procedures would lead to equivalent educational outcomes and to outline the consequence of selecting one testing method over the other.

[1] Direct correspondence to: Dr. Peter Harasym, Division of Educational Planning and Assessment, The Faculty of Medicine, University of Calgary, 2500 University Drive N.W., Calgary, Alberta T2N 1N4

Literature Review

Little empirical research has been reported on the consistency of different standard-setting techniques. Andrew and Hecht (1976) compared the consistency of results using the Nedelsky and Ebel standard-setting techniques. A group of eight judges was selected from a committee which had contributed 180 four-option, type A items to a nationally administered certifying examination. The minimum pass level was set at 68% using the Ebel method and 49% using the Nedelsky method; a 19% discrepancy. The researchers concluded, that the two procedures resulted in significantly different standards despite the fact that both were designed around the same conceptualization of minimum acceptable proficiency. The results suggested that the specific techniques employed in setting an examination standard may be a more powerful determinant of the standard than any other variable. Glass (1978) in discussing this study, reported the educational outcome which would have resulted using each technique (information which he obtained privately):

> *Approximately 95% of the examinees would pass the test if the Nedelsky criterion were used; only 50% would pass the Ebel cut-off (p. 249).*

Methodology

Subject

Two hundred and twelve second year medical students from the graduating classes of 1979, 1980 and 1981 Calgary Medical School took part in this study. The students, enrolled in a three year medical school, had been exposed to a body-systems curriculum designed to teach a "core" body of medical knowledge. The curriculum was organized according to the "mastery" educational philosophy and was taught using a multidesciplinary approach. The students had to complete one more body system before advancing into their clerkship year.

Examination

Three parallel examinations containing type A and type X items were designed to measure student knowledge in the Endocrinology Metabolism (EM) system. The EM course was divided into eight units with each unit under the leadership of a unit manager. The unit manager was responsible for the organization of the unit and the production of multiple-choice items designed to measure the unit's objectives. In addition, managers categorized all items produced. These categorizations led to the setting of a criterion score. After the items were created, a key was produced consisting of the item, the correct answer(s) and the categorizations. The items were reviewed twice: firstly by the EM subcommittee and secondly by a central committee responsible for all evaluations given in the second year. The type A and X items were dispersed throughout the exam according to the unit evaluated. Each exam respectively contained the following numbers of type A items and type X true-false options:Class of 1979: 60 and 126; Class of 1980: 51 and 152; and Class of 1981: 50 and 187.

Standard Setting Procedure

Each unit manager was instructed to make judgements on the options of each item using the following instructions:

I. One-best response (type A)

Identify those options that the "minimumly competent student" must know are incorrect.

II. Multiple true-false (type X)
Identify those options that the "minimumly competent student" must know to be either true or false.

The Nedelsky method was used to determine the mimimum performance level (MPL) for the one-best response (type A) items. Each item was assigned a mark of 1 and an MPL of 1, 0.5, 0.33, 0.25 or 0.20. The criterion score for the test was determined by summing the MPL's for each item. No adjustment factor was added to the criterion score since it represented the consensus of several judges.

For the multiple true-false (type X) items each true-false option was assigned a mark of 1 and the criterion score for the test equalled the number of options the "minimumly competent student" must correctly identify to be either true or false.

Statistical Analysis

Two vectors of dichotomous scores were generated for each student, one for type A responses and one for type X responses. The data for the type A and X items were analyzed separately using the Rasch latent trait model (Wright and Mead, 1977). This statistical procedure enabled the independent calibration of item difficulty and student ability as well as the equating of the scales for the two item types. After calibration, it was possible to identify equivalent scores on the two item-type scales.

Results

The equated mean class abilities (in logits) for item types A and X were found to be equivalent (Class of 1979, 1.95 and 2.07; Class of 1980, 2.29 and 2.34; Class of 1981, 2.28 and 2.29). These results are in keeping with the Rasch model.

The observed MPL's and equivalence and difference in MPL's for Classes 1979-81 on item types A and X are presented in Table 1 below.

insert Table 1

The MPL's are presented in both logits and percentage scores. The observed column contains the observed MPL's for Classes 1979-81 on item types A and X. The equivalent column contains the corresponding MPL on the other item type scale. The different column indicates the discrepancy between the observed and the equivalent MPL of the two scales. The adjustment column represents the calibrating constant for equating the two item-type scales. The arrows represent the cross-over that occurs when the MPL of one item type is mapped onto the MPL of the other item type.

The educational outcomes for the observed and equivalent MPL's are presented in Table 2.

insert Table 2

The table indicates the number of students that were above, below and within one standard error of the observed and equivalent MPL for item types A and X.

Discussion

The results indicated that there were significant discrepancies in MPL's and educational outcomes when students were measured by criterion-referenced tests containing type A and X items. The type A items and the Nedelsky method resulted in MPL's that were consistently lower than that of the type X items (i.e. Class of 1979: 7%; Class of 1980: 9-10%; and Class of 1981: 17-18%). When these MPL's were compared to the student responses on item type A, few students were below the criterion-score for either item type (see observed column of type A items and equivalent scale of type X items in Table 2). However, when the type X item was used, the MPL was consistently higher than the MPL for item type A and a significantly larger number of students were within or below the MPL (see observed column of type X and equivalent column of type A in Table 2).

The findings of this study are in keeping with those reported by Andrew and Hecht (1976). The Nedelsky method of standard-setting led to a significantly lower MPL with differences in educational outcome. Because of the similarity in categorizations made on the options of type A and X items, and the equality of the type A and X scales, it is necessary to support the conclusion of Andrew and Hecht (1976) that specific techniques employed in setting an examination standard may be a more powerful determinant of the standard than any other variable. Thus, the balance of item types and the method selected for setting a standard may be a significant determinant of the educational outcome. Additional research is required to compare other standard-setting procedures and other item types. The Rasch model has proved to be a useful tool by which this may be done.

REFERENCES

Andrew, B.J. and Hecht, J.T. A preliminary investigation of two procedures for setting examination standards. Educational and Psychological Measurement, 36, p. 45-50, 1976.

Burton, N.W. Societal standards. Journal of Educational Measurement, 15, 4, p. 263-73, 1978.

Glass, G.V. Standards and criteria. Journal of Educational Measurement, 15, 4, p. 237-61, 1978.

Hambleton, R.K. Test score validity and standard-setting methods. Criterion-Referenced Measurement: The State of the Art, ed. R.A. Berk. The John Hopkins University Press: Baltimore and London, p. 80-123, 1980.

Hubbard, J.P. Measuring Medical Education: The Tests and the Experience of the National Board of Medical Examiners, Lea & Febiger: Philadelphia, 1978.

Jaeger, R.M. Measurements consequences of selected standard-setting models. Practices & Problems in Competency-Based Measurement, ed. M.A. Bunda and S.R. Sanders, Washington, D.C.: National Council on Measurement in Education, p. 48-58, 1979.

Linn, R.L. Issues of validity in measurement for competency-based programs. In Practises and Problems in Competency-Based Measurement, ed. M.A. Bunda and J.R. Sanders. Washington, D.C.: National Council on Measurement in Education, p. 108-123, 1979.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 43, p. 205-216, 1973.

Messick, S. The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, p. 955-966, 1975.
Shepard, L.A. Measurements consequences of selected standard-setting models. Practices and problems in competency-based Measurements, ed. M.A. Bunda and S.R. Sanders, Washington, D.C.: National Council on Measurement in Education, p. 59-71, 1979.
Wright, B.D. and Mead, R.J. BICAL: Calibrating items and scales with the Rasch model, Statistical Laboratory, Dept. of Education, The University of Chicago, Research memorandum 23, Jan., 1977.

Table 2

Educational Outcome

Class	MPL	Type A O*	Type A E**	Type X O	Type X E
1979 n=70	above	66	42	49	68
	within	2	14	19	2
	below	2	4	2	0
1980 n=70	above	61	53	49	62
	within	9	12	12	8
	below	0	5	9	0
1981 n=71	above	70	40	52	71
	within	1	25	12	0
	below	0	6	7	0

O* observed
E** equivalent

Table 1

Minimum Performance Level

Class	Type	observed logits	observed # score	equivalent logits	equivalent % score	difference % score	adjustment logits
1979	A	0.29 ± 0.29	56%	0.864 ± 0.22	65%	7%	-.406
	X	1.27 ± 0.23	72%	0.696 ± 0.30	63%	7%	
1980	A	0.36 ± 0.32	58%	1.24 ± 0.22	72%	9%	.605
	X	1.84 ± 0.25	81%	0.97 ± 0.34	68%	10%	
1981	A	0.51 ± 0.36	59%	1.24 ± 0.22	76%	17%	.179
	X	1.75 ± 0.23	79%	0.69 ± 0.36	61%	18%	

A COMPARISON OF SEVERAL SCORE CUTTING PROCEDURES AND THEIR EFFECTS ON SUCCESS RATES

Ernest N. Skakun* Sam Kling

R.S. McLaughlin
Examination and Research Centre,
Edmonton Alberta

INTRODUCTION

One of the problems facing institutions and agencies responsible for licensure and certification of medical trainees is that of establishing a cutting or passing score on an examination. Traditionally, the strategy used to establish a cutting score was to resort to custom and follow the practice of other institutions or that offered by legislation. Thus, cutting scores were set at a fixed percentage (for example, 50%), or they were set at a specified distance in standard deviation units from the mean of the test scores (for example, one standard deviation below the mean). Such practice remained virtually unchallenged until the minimum competency testing movement appeared on the assessment scene. The minimum competency testing movement reaffirmed the need for standards.

A variety of procedures for setting cutting scores have been described. Very roughly, these procedures can be categorized as those based on test content (Nedelsky, 1954; Angoff, 1971,; Jaegar, 1976; Ebel, 1979); those based on group performance (Zieky and Livingston, 1977) and those based on empirical methods (Block, 1972; Kriewall, 1972; Hambleton and Novick, 1973; Berk, 1976; Livingston, 1976; Huynh, 1976; Van der Linden and Mellenbergh, 1977; Schoon, Gullion and Ferrara, 1979).

Reviews of these procedures have been provided by Millman (1973), Meskauskas (1976), Glass (1978), Glass and Smith (1978), Hambleton, Powell and Eigor (1979) and Shepard (1980). According to Hambleton et al. (1979), most of the work on cutting scores has been on the suggestion of methods, rather than on actual empirical investigation of methods. Work of the empirical nature has been conducted by Andrew and Hecht (1976), Meskauskas and Webster (1975), Brennan and Lockwood (1979), and Paiva and Vu (1979).

The purpose served by the present study is to compare several procedures for establishing cutting scores. More specifically, the study investigates whether the Nedelsky, two modified Ebel procedures and norm-referenced approaches, generate similar cutting scores. Lastly the study investigates the effects of the cutting scores derived from the various methods on the overall pass rate.

REVIEW OF THE LITERATURE

Andrew and Hecht (1976) used a panel of eight judges, who met on two separate occasions, to set a passing score using the Nedelsky and Ebel procedures. The percentage of items expected to be answered correctly by the minimally qualifiable candidate was 68% for the Ebel procedure and 49% by the Nedelsky approach. Glass (1978) reported that this 20% difference would fail 50% of the candidates according to the Ebel criterion, and 5% of

* Direct correspondence to: Ernest N. Skakun, Assistant Director, R.S. McLaughlin Examination and Research Centre, Room 222, Campus Tower, 8625, University of Edmonton, Edmonton, Alberta, T6G 1K9, Canada.

the candidates would have failed had the cutting score derived from the Nedelsky approach been implemented. Andrew and Hecht attributed the difference to different philosophical assumptions and varying conceptualizations, while Hambleton (1978) stated that because of the difference in procedures and instructions, one would expect the cutting scores derived from the two procedures to be different. Brennan and Lockwood (1979) used generalizability theory to quantify the variability produced in the cutting scores derived from the Nedelsky and Angoff procedures. The Nedelsky procedure produced lower cut-off scores and greater variability in judges' ratings. In addition, the variance components estimated from a mixed-effects ANOVA were four times greater for differences in procedures than for differences in rater means. The variability of the judges' ratings in the Nedelsky procedure were similar to those reported by Meskauskas and Webster (1975). The range for cutting scores was from 36% to 80% for single-best-answer questions, and from 48% to 89% for true/false questions. Paiva and Vu (1979) suggested that the great difference in cutting scores produced by the judges using the Nedelsky technique could be attributed to the judges' difficulty in dissociating their judgments from their own difficulty in answering the questions.

METHODOLOGY

In the Nedelsky approach, judges are instructed to review each multiple choice question and decide how many of the options a barely qualifiable examinee would be able to reject as being obviously wrong. The cutting score for each item is determined by computing the chance score for the remaining options. Each judge's cutting score is obtained by summing the chance scores across all items. A cutting score is obtained by arranging all the judge's cutting scores. Nedelsky felt that by computing the standard deviation of the individual judge's cutting scores, the distribution would be synonymous with the hypothesized distribution of the scores of the barely qualifiable examinees. This standard deviation would then be multiplied by a constant and the average cutting score adjusted upwards or downwards by this amount. In the present study, the use of the standard deviation and constant term were ignored.

The procedure proposed by Ebel is more complex. Judges are asked to rate items along two dimensions - relevance and difficulty. Ebel uses four categories of relevance - essential, important, acceptable and questionable - and three difficulty levels - easy, medium and hard. These categories form a 3 x 4 grid and judges are asked to locate each of the test items in the proper cell of the grid and secondly to assign to each cell a number representing the proportion of items in that cell that a barely qualifiable examinee should be able to answer. The number of items in each cell is multiplied by the appropriate proportion and the sum of the products yield a cutting score for each judge. The final cutting score is the average of the individual judge's decisions.

For the purpose of the present study, several modifications were made to the Ebel procedure. Instead of using a relevance by difficulty grid, items were categorized according to a difficulty by taxonomy (DT) grid and a relevance by taxonomy (RT) grid. Three categories of relevance - essential, important and acceptable - three levels of difficulty - easy, medium and hard - and three categories of taxonomy - factual, comprehension and problem solving - were used in the present study.

In the traditional Ebel method, judges are first asked to classify the items appropriately into one of the cells formed by the relevance by difficulty grid, and secondly to indicate what proportion of the questions falling into each cell of the grid a barely qualifiable candidate is expected to answer correctly. Since the items used in the present study were already classified according to taxonomy and relevance, and since they had all been administered at least once, the only task for the judges in the present study was to indicate for each cell the proportion of questions a barely qualifiable candidate should be expected to answer correctly.

The examination used in the present study consisted of 194 multiple choice questions which were part of a nationally administered examination. This examination was administered to 168 candidates seeking certification in the medical specialty of General Surgery. The items all originated from the General Surgery test item library, where each item is classified according to difficulty, discrimination, content area, taxonomic level and relevance. The policy of the General Surgery Examining Board is to set the passing score at one standard deviation below the mean performance of the candidates comprising the Reference Group. Reference Group candidates are defined as those candidates who have graduated from approved North American medical schools and who are writing the examination for the first time. All other candidates constitute the Non-Reference Group. At the time that the investigation was conducted, 48 candidates comprised the Reference Group and 120 comprised the Non-Reference Group. Thus, one of the norm-referenced approaches compared was the one establishing the cutting score at one standard deviation below the mean. The second norm-referenced approach consisted of averaging the cutting scores derived from the criterion of one standard deviation below the mean, which was established each time the national examination was administered in the last five years.

Eight judges actively involved in the General Surgery Test Committee with the writing of multiple choice items and test preparation participated in the study. Of the eight, seven participated using the Nedelsky method and six judges participated in each of the two modified Ebel methods. The Nedelsky procedure was completed first. This was followed by a six month interval, at the end of which six judges used the taxonomy by difficulty grid to establish a cutting score. Three days later, the same six judges participated in establishing a cutting score using the relevance by taxonomy grid.

Passing scores were determined by each judge for the Nedelsky and each of the modified Ebel procedures. The final cutting score for each method was determined by averaging the judges' scores. The information was summarized, using descriptive statistics and the number of failures resulting from each approach was tabulated. A reliability of the ratings given to the items for the Nedelsky and two modified Ebel approaches was estimated, using analysis of variance.

RESULTS

Table 1 presents the descriptive statistics for the various methods.

Table 1 Examination Standards & Descriptive Statistics

	Nedelsky	DT	RT
Number of judges	7	6	6
Range	.477 - .781	.613 - .757	.644 - .784
Mean	.667	.697	.717
Standard Deviation	.093	.056	.058
Cutting score	129	135	139

The passing scores range from 129 items (66.7%) for the Nedelsky technique to 135 (69.7%) and 139 (71.7%) for the two modified Ebel methods. The cutting score using the criterion of one standard deviation below the mean was established at 137 items (70.6%), while the cutting score based on averaging the cutting scores over the last five years was 68.5% (132 items) with a standard deviation of 0.85.

The effects of the passing scores determined by the various approaches are presented in Table 2 for Reference and Non-Reference Group candidates:

Table 2 Number of Failures according to Method and Group.

	Nedelsky	DT	RT	-1S	Average
Ref. Group (48)	3	5	7	6	4
Non-Ref. Group (120)	33	51	66	60	44
Total Group (168)	36	56	73	66	48
% of Total Group	22.5	35.0	45.6	41.3	28.5

Depending upon which approach is uscd, a different number of candidates fail. Slightly more than 20% of the examinees would fail had the Nedelsky technique been employed. The modified Ebel approach based on a categorization of items according to difficulty and taxonomy would fail 35%, while the approach using relevance and taxonomy would fail 45.6% of the candidates. This is contrasted with the failure rate of 41.3%, derived from establishing the cutting score at one standard deviation below the mean. The number of failures based on the average of the cutting scores over the past five years would fail 48 or 28.5% of the candidates.

The reliability of the ratings given by the seven judges using the Nedelsky procedure was 0.61. This is compared with reliability estimates of 0.98 for each of the two modified Ebel approaches.

DISCUSSION

Clearly, the methods produce different cutting scores. However, the differences are not as great as those reported by Andrew and Hecht (1976). The Nedelsky method produced the lowest cutting score, while the modified Ebel method based on a classification of items according to relevance and taxonomy produced the highest cutting score. The difference between these two is in the order of 5%. The cutting scores produced by the norm-referenced approaches fall in between. Shepard (1980) states that between the Nedelsky and Ebel procedures, the Nedelsky method will produce the lower cutting score, since the task of eliminating clearly wrong answers is easier than choosing the correct answers and indicating what proportion of the candidate population would be expected to answer the questions correctly. The Nedelsky procedure also restricts the judges' decisions to a small number of unequally spaced probabilities. The Nedelsky procedure could also be questioned on the basis of whether the task for the judges is consistent with how a barely qualifiable candidate would answer the question. All of these factors might contribute to the modest reliability estimate obtained for the judges' rating of the items, and to the variability of cutting scores produced by the individual judges.

As great a variability of judges' cutting scores was not witnessed for the two modified Ebel approaches, part of the variability was reduced by presenting a single task for the judges, namely, an indication of what proportion of the items should be answered correctly. The cutting score established using the criterion of one standard deviation below the mean was set at 70.6% and this represented the highest cutting score in the five year period. Although the criterion remained the same, the cutting scores during the five year period ranged from 65.6% to 70.6%.

Although the difference between the cutting scores set by the Nedelsky and the second modified Ebel approach represents only 10 items or 5%, this difference has a more pronounced effect on the overall success rate. Had the Nedelsky criterion been implemented, 36 candidates (22.5%) would have failed. On the other hand, 73 candidates (45.6%) would have failed according to the modified Ebel approach. This is contrasted with the actual failure rate of 41.3% (66 candidates) determined by the criterion of one standard deviation below the mean performance of the Reference Groups.

From the standpoint of methodology, the present study could be criticized for using a small number of judges, for not providing a more detailed definition of the barely qualifiable candidate, and for the small time interval between the judges' participation in the two modified Ebel methods.

CONCLUSIONS AND IMPLICATIONS

One of the problems associated with determining a cutting score is that a continuously distributed trait must be artificially dichotomized to produce a pass-fail line. This pass-fail line is represented by a single point on the continuum and is taken as the distinction between the competent and incompetent, suggesting that competency is achieved by crossing a threshhold or barrier. Such a view leads to arbitrariness in setting a cutting score and difficulty in choosing an approach from the many available procedures. In the present study, it was demonstrated, although with limited generalizability, that different approaches generate similar cutting scores. However, a difference of 5% in the scores can produce drastic effects on overall success rates. In keeping with the suggestions made by Shepard (1980) it is recommended that a combination of different approaches using a large number of judges representing important audiences be used.

REFERENCES

Andrew, B.J., & Hecht, J.T. A preliminary investigation of two procedures for setting examination standards. Educational and Psychological Measurement, 1976, 36, 35-50.

Angoff, W.H. Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.) Educational Measurement, Washington, D.C.: American Council on Education, 1971.

Berk, R.A. Determination of optimal cutting scores in criterion-referenced measurement. Journal of Experimental Education, 1976, 43, 4-9.

Block, J.H. Student learning and the setting of mastery performance standards. Educational Horizons, 1972, 50, 183-190.

Brennan, R.L., and Lockwood, R.E. A comparison of two cutting score procedures using generalizability theory. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1979.

Ebel, R.L. Essentials of Educational Measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1979.

Glass, G.V. Standards and criteria. Journal of Educational Measurement, 1978, 15, 237-261.

Glass, G.V., and Smith, M.L. The Technology and Politics of Standards. Educational Technology, 12-18, 1978.

Hambleton, R.K. On the use of cut-off scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 1978, 15, 277-290.

Hambleton, R.K., and Novick, M.R. Towards an integration of theory and method for criterion referenced tests. Journal of Educational Measurement, 1973, 10, 159-170.

Hambleton, R.K., Powell, S., and Eignor, D.R. Issues and methods for standard setting. In A practitioner's guide to criterion-referenced test development, validation and test score usage. Laboratory of Psychometric and Evaluative Research Report No. 70. Amherst, Mass.: School of Education, University of Massachusetts, Amherst, 1979. (2nd ed., 480 pps).

Huynh, H. On the reliability of decisions in domain-referenced testing. Journal of Educational Measurement, 1976, 13, 253-264.

Jaeger, R. Measurement consequences of selected standard-setting models. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, April 1976.

Kriewall, T.E. Aspects and applications of criterion-referenced tests. Paper presented at the annual meeting of AERA, Chicago, 1972.

Livingston, S.A. A utility-based approach to the evaluation of pass/fail testing decision procedures. Report No. COPA-75-01. Princeton, N.J.: Center for Occupational and Professional Assessment, Educational Testing Service, 1975.

Meskauskas, J.A. Evaluation models for criterion-referenced testing: Views regarding mastery and standard-setting. Review of Educational Research, 1976, 45, 133-158.

Meskauskas, J.A., and Webster, G.D. The American Board of Internal Medicine Recertification Examination: Process and Results. Annals of Internal Medicine, 1975, 82, 577-581.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 43, 205-216.

Nedelsky, L. Absolute grading standards for objective tests. Educational and Psychological Measurement, 1954, 14, 3-19.

Paiva, R.E.A., and Vu, N.V. Standards for acceptable level of performance in an objectives-based medical curriculum: A case study. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1979.

Schoon, C.G., Gullion, C.M., and Ferrara, P. Bayesian Statistics, Credentialling Examinations, and the Determination of Passing Points. Evaluation and the Health Professions, 2:2, 181-201, 1979.

Shepard, L.A. Technical Issues in Minimum Competency Testing in Review of Research in Education, Volume 8, Itasca ILL: F.E. Peacock, Publishers, 1980.

Van der Linden, W.J., and Mellenbergh, G.J. Optimal cutting scores using a linear loss function. Applied Psychological Measurement, 1977, 1, 593-599.

Zieky, M.J., and Livingston, S.A. Manual for setting standards on the Basic Skills Assessment Test. Princeton, N.J.: Educational Testing Service, 1977.

Content Representativeness and Student Performance on National Board Part I Special Subject Examinations*

William C. McGaghie, Office of Medical Studies
Hugh J. Burford, Department of Pharmacology
Donna H. Harward, Office of Medical Studies

University of North Carolina at Chapel Hill

Freshman and sophomore medical students at the University of North Carolina (UNC) must pass end-of-year comprehensive examinations for promotion. Each subject in the medical curriculum is covered and each year-end examination is composed of selected special subject examinations (subtests) from the National Board of Medical Examiners (NBME) Part I test as well as locally prepared tests. On approval from a faculty committee having responsibility for basic science examinations, UNC course directors may opt to use an NBME special subject examination in lieu of a local test.

During the 1978-79 academic year, seven of 23 basic science course directors chose to use one or more NBME special subject examinations to evaluate student learning in their courses. The courses included both departmental offerings (e.g., Pharmacology) and interdisciplinary courses taught from an organ system format (e.g., Endocrine System). Five National Board Part I subtests (Anatomy, Pharmacology, Physiology, Pathology, and Microbiology) were employed in combination to measure student achievement in the seven courses. Table 1 displays the distribution of the five Part I subtests among the seven basic science courses.

The NBME special subject examinations are secure tests. Specific features of their architecture and the medical content sampled by the examination questions are closely guarded secrets. Faculty inspection of the special subject examinations is allowed only after the exams have been administered to medical students, scored, and the results reported to local decision makers.

Given the security of the NBME special subject examinations, medical faculty make a key assumption when they use the external tests to evaluate intramural student learning. The faculty members assume that the content of the tests accounts for a representative sample of the curricular content presented to students. (It is important to note that the representativeness of test content is a necessary but not a sufficient condition toward the achievement of content validity. As Messick [1] and Guion [2] have argued, content validity is not only a function of test stimuli but also of examinee responses.) There is, in addition, a corollary to the content representativeness assumption, _viz._, that questions on content that has received much local emphasis will be answered correctly by a larger proportion of medical students than questions on material receiving less educational emphasis. The primary assumption concerns the match between special subject examination questions and local educational objectives. Its corollary presumes a correlation between teaching emphases and student examination performance.

*Address reprint requests to: Office of Medical Studies, University of North Carolina School of Medicine, 322 MacNider Hall 202H, Chapel Hill, N.C. 27514

Table 1

Distribution of NBME Subtests Among UNC Basic Science Courses

UNC Courses	NBME Part I Subtests					Tot. NBME Exams Per Course
	Anat	Pharm	Phys	Path	Micro	
1	X					1
2	X					1
3		X				1
4		X	X	X		3
5		X	X	X		3
6			X			1
7					X	1
Tot. Courses Per NBME Exam	2	3	3	2	1	

This study is designed to test the accuracy of the assumption and its corollary at the UNC School of Medicine. Specifically, the degree to which the item content of five NBME Part I examinations received instructional emphasis is assessed for the 1978-79 academic year. Correlations between faculty judgments of the instructional emphasis given to the content of NBME exam items and student performance on those items are subsequently presented. The report concludes with a discussion of several implications the research holds for medical education at UNC and in other settings.

Related Literature

The 1979-80 AAMC Curriculum Directory indicates that the use of external National Board examinations to reach internal decisions about medical students is widespread among American medical schools. For example, 57 of 125 U.S. medical schools (45.6%) require their students to pass the Part I examination for promotion. In addition, the NBME notes in its 1978 Annual Report (3) that, "At the present time, performance on National Board examinations plays some part in academic decision making in approximately 80 percent of all U.S. medical schools" (p. 9). However, in the same report the NBME issues several warnings to medical schools that use its examinations to reach internal decisions about students: (a) the primary purpose of NBME examinations is to provide data needed to inform certification (and licensure) decisions, not educational decisions; (b) medical faculty tend to attribute far more meaning to NBME test scores than is warranted; (c) the examinations are designed to measure cognitive knowledge which is only one facet of medical competence; and (d) "Within the cognitive domain, an externally developed examination will inevitably include content areas not addressed in a given curriculum, and, in turn, a given curriculum may include content not addressed in a given examination" (p. 10).

The last warning given by the NBME speaks directly to the content representativeness of its examinations and indirectly to their utility for reaching educational decisions at individual medical schools. Several studies have addressed this subject.

Kennedy, Kelley, and Hubbard (4) surveyed U.S. and Canadian medical schools on behalf of the NBME to determine the "relevance of Part I National Board examinations in basic medical sciences to current medical school curricula" (p. 5). Participation was obtained from 103 schools where deans and faculty members reviewed samples of basic science test items drawn from a larger item pool. Two questions were posed for each item: (a) "Would this item be appropriate for inclusion in an examination to be given by your department to your students at the end of the required course in this subject? (Yes or No);" and (b) "If the answer to question A concerning this item is 'yes,' should it be included in a specially selected group of items on which each student should be required to perform very well (i.e., no fewer than 75% percent correct responses in the selected group as a whole) in order to achieve a minimal 'pass' in the course? (Yes or No)" (p. 5). The results showed that across six basic science subjects, 90% of approximately 1800 test items were judged appropriate for inclusion in basic science course examinations. Another finding from the NBME survey was that 65% of the items were not only deemed appropriate, but also essential for those courses.

Investigators at individual medical schools have also studied the match between the National Board Part I examination and their local curricula. Garrard, McCollister, and Harris (5) investigated the relationship between Part I test items and the content of courses at the University of Minnesota Medical School. Basic science course coordinators rated each item in six subtests according to the degree of teaching emphasis its content received. The ratings were made using a 4-point scale ranging from maximum emphasis to no emphasis. Results from the study showed that, "Course content receiving moderate or maximum teaching emphasis during the first year alone were consistent with approximately 85% of the anatomy and microbiology subtest questions and 70% of those in biochemistry. Combining the first two years of the curriculum, material given moderate or maximum emphasis covered 86% of the subtest questions in pharmacology, 83% in pathology, and 65% in physiology" (pp. 424-425).

In an analogous study at the Case Western Reserve University (CWRU) School of Medicine, Wile (6) solicited faculty judgments about the curricular relevance of the items contained in seven NBME Part I subtests. For each relevant item, faculty reviewers were also asked to indicate whether the information needed to answer the question was or was not taught by local subject committees, or if the reviewers did not know if the information was presented. Wile's findings show that across the seven subtests, 85% of the test items were both relevant to the CWRU basic science curriculum and taught by its faculty. Of the remaining items, 12.2% were judged to have curricular relevance although the information needed to answer them was not presented; 2.7% of the items were deemed not relevant to the CWRU basic science program.

Largely based on the results of this study, the CWRU faculty no longer requires medical students to take the NBME Part I examination. Wile (6) reports that the faculty reasoned, ". . . much of what was asked was not taught in the second-year curriculum and significant areas taught in Phase 2 were not tested by this examination" (p. 96). The mismatch between Part I content and curricular content was cited for inappropriate student attention to the examination rather than to coursework. Further, "As a consequence of this study, the faculty recognized that the NBME Part I examination was appropriate for licensure but inappropriate for internal student evaluation" (p. 96).

The varied findings from these studies suggest that individual medical schools should carefully review the NBME Part I examination to determine if the exam item content coincides with curricular content. The available evidence indicates the content representativeness assumption may be supported, yet it also may not.

But do medical students perform better on test items drawn from content the faculty has emphasized in teaching? A preliminary answer would be affirmative because both Garrard et al. (5) and Wile (6) present comparative (t-test) evidence which shows that medical students perform significantly better on test items whose content has been presented by their teachers. The present study is similar, but not an exact replication, of these two previous investigations. Rather than assessing group differences (e.g., freshman vs. sophomore) in terms of test items placed in different categories (e.g., relevant and taught, relevant and not taught), this study evaluates course-specific relationships between faculty ratings of teaching emphasis given to the content of NBME items and student performance on those items.

Methodology

Review copies and item analyses of five NBME Part I subtests (Anatomy, Pharmacology, Physiology, Pathology, Microbiology) were obtained from the National Board following their administration as components of UNC freshman and sophomore comprehensive examinations in the Spring, 1979. The seven basic science course directors who used the special subject examinations to represent the content of their courses rated each NBME subtest item using a modified version of the scale reported by Garrard et al. (5). The course directors were asked to rate "the degree to which information needed to answer each question was emphasized in your course or clerkship." Scale values ranged from zero (no emphasis) to three (maximum emphasis) with two intermediate levels (minimum and moderate emphasis). The instructions asked faculty raters to be attentive not only to different levels of stress in formal lectures, but also to stress in assigned readings and other educational media.

The data were analyzed in two stages. First, the number and percentage of NBME subtest items receiving instructional emphasis (rating of 1, 2, or 3) was tabulated for each basic science course. Items on content that received emphasis in more than one course were omitted to prevent item-course overlap. These tabulations describe the match between internal teaching goals and the content representativeness of the external tests. Second, nonparametric correlation coefficients (Kendall's tau) were calculated between (a) emphasis ratings of items <u>unique</u> to individual courses, and (b) the proportion of UNC medical students who correctly answered each question (local item difficulty index or p-value). A nonparametric analysis was performed because it could not be safely assumed that the emphasis ratings met the measurement assumptions needed for Pearson correlations (7). Here, subtest items are the unit of analysis and items rated zero (no emphasis) were excluded. These correlations describe the relationship between course directors' perceptions of content emphasis and student performance on test items representing that content.

Results

Results from the two analyses are summarized in Table 2. The first analysis assessed the congruence of local teaching goals and the item content of the NBME subtests. Inspection of the entries contained in the bottom row of the table indicates much variation in the match between test content and instructional emphases for the seven basic science courses. The percentage of NBME subtest items having content that received at least minimum teaching

Table 2

Content Representativeness (E Items) and Correlations (Kendall's τ) Between NBME Item p-Values and Faculty Ratings of Item Instructional Emphasis

UNC Courses	NBME Subtests									
	Anatomy (136 Items)		Pharmacology (142 Items)		Physiology (135 Items)		Pathology (138 Items)		Microbiology (135 Items)	
	E Items	τ	E Items	τ	E Items	τ	E Items	τ	E Items	τ
1	55	.04								
2	51	.34**								
3			110	.15*						
4			6	.54	21	.55**	14	.69**		
5			--	--	10	.03	--	--		
6					18	-.44*				
7									112	.16**
No. and % of Unique Items Receiving Instructional Emphasis	106 (78%)		116 (at least)[a]		49 (36%)		14 (at least)[a]		112 (83%)	

Note. Median τ=.16. Course No. 5 also used the Pharmacology and Pathology subtests although item emphasis ratings were not provided.

[a] Meaningful percentages cannot be calculated due to missing data.

*$p \leq .05$

**$p \leq .01$

emphasis ranges from 36% for the Physiology subtest to 83% for the Microbiology subtest. Calculation of similar percentages for the Pharmacology and Pathology subtests is not possible due to missing data from Course No. 5. It is important to stress that only items receiving instructional emphasis unique to individual courses are included in these figures. However, few items were lost by this procedure because overlapping (i.e., redundant) emphasis ratings were assigned by the course directors to only one item for the Anatomy subtest and three items from the Pharmacology subtest. No overlapping emphasis ratings were obtained for items contained in the Physiology, Pathology and Microbiology subtests.

The results from the correlational analysis involving course directors' ratings of item content emphasis and item p-values are given in the upper portion of Table 2. The nine coefficients are widely divergent in magnitude, direction, and the number of test items involved in their calculation. The strongest positive correlation (.69) is observed for Pathology test items (n=14) whose emphasis was rated by the director of Course No. 4. Such a coefficient indicates that the course director has a clear view of the elements of Pathology that received varied emphasis in the course and that student learning is strongly associated with that subset of teaching goals. A contrary result is seen for the Physiology subtest within Course No. 6. Here, a pronounced negative correlation (-.44, n=18) suggests that perceived course emphases are inversely proportional to medical student learning. The median of the nine correlations, across five NBME subtests and seven UNC basic science courses, is .16.

Discussion

The data presented in this report suggest that the content representativeness of the five NBME special subject examinations that have been used as measures of student achievement at the UNC School of Medicine has not been firmly established. It has been shown from the judgments of the basic science course directors who opt to use the external examinations that the match between teaching emphases and test content coverage varies widely. This finding differs from those cited earlier (4-6) where closer correspondence was observed between the aims of medical curricula and the content of extramural medical examinations. In addition, it underscores statements by Cronbach (8) that the content coverage of nationally standardized tests should be carefully assessed by local evaluators. Cronbach argues that content representativeness must be demonstrated, not assumed, to insure accurate interpretation of test results at individual educational institutions.

Much variation was also observed in the relationships between course directors' ratings of the educational emphasis given to the content of NBME subtest items and student performance on those items. There are several plausible explanations for this result. For example, the courses represented in this study were taught by multiple instructors. Course directors who were unfamiliar with the presentations and assignments of their teaching colleagues may have provided emphasis ratings from an incomplete information base. Another possible explanation is that a discrepancy exists between faculty beliefs about the stress given to different bodies of course content and student perceptions of content emphasis. The inability to achieve consistently high and positive correlations between faculty judgments of content emphasis and student performance suggests a need for local efforts to tighten the fit between teaching and testing.

Conclusions and Implications

Medical schools that use external tests as internal measures of student learning assume the content coverage of the tests is a representative sample of the content presented in the curriculum. This study indicates that such as assumption should be tested at the local level before its accuracy is established. The content representativeness of NBME special subject examinations is not a sure bet. It needs to be demonstrated by each medical school that uses the tests to evaluate medical students.

References

1. Messick, S. The Standard Problem: Meaning and Values in Measurement and Evaluation. Am. Psychologist, 30:955-966, 1975.

2. Guion, R.M. Content Validity — The Source of My Discontent. Applied Psychol. Meas., 1:1-10, 1977.

3. National Board of Medical Examiners. 1978 Annual Report. Philadelphia: Author, 1979.

4. Kennedy, W.B., Kelley, P.R., and Hubbard, J.P. The Relevance of National Board Part I Examinations to Medical School Curricula. Philadelphia: National Board of Medical Examiners, 1970.

5. Garrard, J., McCollister, R.J., and Harris, I. Review by Medical Teachers of a Certification Examination: Rationale, Method and Application. Med. Educ., 12:421-426, 1978.

6. Wile, M.Z. External Examinations for Internal Evaluation: The National Board Part I Test as a Case. J. Med. Educ., 53:92-97, 1978.

7. Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

8. Cronbach, L.J. Test Validation. In Educational Measurement (2nd ed.). R.L. Thorndike (Ed.). Washington, D.C.: American Council on Education, 1971.

THE FEASIBILITY OF USING A CANADIAN EXAMINATION FOR THE CERTIFICATION OF AUSTRALASIAN CANDIDATES IN PSYCHIATRY

Ernest N. Skakun[1]
R.S. McLaughlin Examination
and Research Centre

Donald R. Wilson
Gladys and Merrill
Muttart Foundation

Introduction

Agencies and institutions responsible for the licensure and certification of medical trainees are common to many nations. Thus, in the United States the National Board of Medical Examiners and the various specialty Boards fulfill this function. Comparable to these are the Medical Council of Canada and the Royal College of Physicians and Surgeons of Canada, the Institute for Research in Medical Education and Evaluation in Switzerland, and the various Royal Colleges in Great Britain, Australia, and New Zealand. No doubt, there is a great exchange of ideas regarding research and development between these institutions, and one of the cooperative ventures to arise from such exchanges is the international sharing of test material. Published studies of such efforts are few; the most recent involves the National Board of Medical Examiners (NBME) and the Swiss Institute for Research in Medical Education and Evaluation (IAE). A test developed by the NBME and administered to American medical students was translated into French and German and administered to a group of Swiss medical students as a graduating examination (Burri, Schumacher, and Vorkauf, 1977). The authors concluded that the translated examination was a valid test for Swiss students.

Along somewhat similar lines of endeavor, the Royal College of Physicians and Surgeons of Canada and the Royal Australian and New Zealand College of Psychiatrists entertained an agreement whereby a Canadian developed examination would be reviewed by the Australasian College and administered to candidates in Australia and New Zealand. The purpose served by the present study is to investigate the feasibility of such an endeavor. More specifically, the study reviews the selection of items, compares the performance of Canadian and Australasian candidates on the common examination, and compares the results of the item analyses.

[1]Direct correspondence to: Mr. Ernest N. Skakun, Associate Professor and Assistant Director, The Royal College R.S. McLaughlin Examination and Research Centre, University of Alberta, 222 Campus Towers, Edmonton, Alberta, T6G 1K9.

Methodology

The certification process of the Royal College of Physicians and Surgeons of Canada requires candidates seeking certification in the specialty of Psychiatry to take a 240 item multiple choice examination. The examination administered in 1977 to Canadian candidates served as the source of items that were reviewed and selected by the Royal Australian and New Zealand College of Psychiatrists. All 240 items had been reviewed by a Canadian test committee for relevance, structure, and correct answer. The 240 items were then reviewed by the Australasian panel of content experts for relevance. Of the 240 items, 160 were selected for administration. Items that were rejected were done so for geographic reasons. For example, laws regarding mental health in Canada are not applicable in Australia and New Zealand. The 160 selected items were administered at three different testing times during 1977 and 1978 to Australasian candidates. Once the examinations were written, they were scored and the questions submitted to an item analysis. All analyses treated the Australasian candidates as three separate groups. Descriptive statistics, including mean performance, dispersion, standard error of measurement, and reliability estimates (KR-20) were computed for each test administration. Item analysis based on each administration was also performed.

Results and Discussion

Table 1 presents the descriptive statistics.

	Canadian Candidates	Australasian Candidates 1	Australasian Candidates 2	Australasian Candidates 3
Number of candidates	63	32	29	25
Mean	108	98	100	106
Standard deviation	12	15	14	9
Standard error of measurement	5	5	5	5
Reliability (KR-20)	0.84	0.91	0.89	0.72
Range of scores	79-138	56-125	64-121	86-121
Number of items	160	160	160	160

Table 1: Descriptive Statistics

Results of these comparisons indicate that Canadian and Australasian candidates seeking certification in Psychiatry do equally well on examinations. If we compare Australasian performance with the performance of the Canadian group, then in the early administrations, the Australasian candidates had a lower mean score, for example, 98 (Time 1, March, 1977), 100 (Time 2, August, 1977), and 106 (Time 3, August, 1978) compared to 108 for the Canadian group. The difference of 10 items, or about 6%, bears no significant or practical difference. A review of

35

performance across subspecialties revealed no significant difference between the Canadian group and the Australasian candidates.

Turning to the item analyses, Table 2 presents the distribution of the item difficulty indices. Difficulty as used here refers to the proportion or percentage of candidates answering the question correctly.

		Australasian		
	Canadian	1	2	3
<30	13	21	19	18
30 - 79	80	80	89	74
≥80	67	59	52	68

Table 2: Distribution of Difficulty

The number of difficulty items appearing in the test is small, both for the Canadian and Australasian candidates. Approximately 20 items out of the 160 are answered correctly by no more than 30% of the candidate population. On the other hand, there is a preponderance of easy items. About 60 out of the 160 items are answered correctly by at least 80% of the candidate population. In general, the difficult items were difficult for all candidate groups. Likewise, items that were easy were easy for all candidate groups as well.

Table 3 shows the results of dividing the items into three groups of difficulty. Group 1 were those items that were easier for Canadian candidates than for Australasian candidates. A difference of 10% in the difficulty index was set as an arbitrary measure. Group 2 were those items where the difference was within the 10% range, and Group 3 consisted of those items that were easier for the Australasian candidates. Once again, the 10% difference was used.

	Australasian		
	1	2	3
1. Canadian-Australasian >10%	47	41	24
2. Australasian-Canadian <10%	102	102	108
3. Australasian-Canadian >10%	11	17	28

Table 3: Ease of Items

On the basis of the first test administration, 47 items were identified as being easier for Canadian candidates, 102 items were identified as showing no difference in performance, and 11 items were identified as being easier for Australasian candidates. For the second and third administration, the number of items that were identified as being easier for Canadian candidates drops, while the number of items identified as being easier for Australasian candidates increases. The 10% difference is a stringent rule, considering that the difference of 10% can be brought about by having 2 or 3 Australasian candidates answer the question incorrectly or correctly.

Table 4 presents the distribution of discrimination indices.

	Canadian	Australasian 1	Australasian 2	Australasian 3
<.20	78	72	72	100
.20 - .29	36	23	34	23
.30 - .39	29	29	27	18
.40 - .49	16	19	12	12
>.50	1	17	15	7

Table 4: Distribution of Item Discrimination Indices

Generally, an item having a discrimination index of less than 0.20 is considered as a poor indicator of examinee differentiation. Thus, 78 out of the 160 items did not meet this criterion of 0.20 or greater using the Canadian group as a basis for analysis. The number of non-discriminatory items using the Australasian candidates as a basis for analysis reveals that there were 72 non-discriminatory items in test administration 1 and 2, and 100 items in test administration 3. At the other end of the scale, item analysis based on the Australasian candidates revealed a higher number of discriminatory items. Thus, there was only one Canadian item with a discrimination index of .50 or greater, compared to 17, 15, and 7 such items for the Australasian candidates.

Table 5 presents a cross tabulation of the discrimination indices.

		Australasian 1		Australasian 2		Australasian 3	
		<0.20	≥0.20	<0.20	≥0.20	<0.20	≥0.20
Canadian	<0.20	45	33	40	38	47	31
	≥0.20	27	55	32	50	53	29

Table 5: Comparison of Discrimination Index for Canadian and Australasian Candidates

37

Thus, from the first test administration, there were 45 items that were identified as weak discriminators for both Canadian and Australasian candidates, 27 items that were discriminatory for the Canadian candidates but not for the Australasian candidates, 33 that were non-discriminatory for the Canadian candidates but discriminatory for the Australasian candidates, and 55 items that were discriminatory for both groups of candidates. The rest of Table 5 presents similar information for Australasian test administrations 2 and 3.

From the methodological view several criticisms could be aired regarding the study. The three groups of Australasian candidates could have been grouped together. In doing so, it can be speculated that the results would be similar to those reported. However, it was chosen to analyze the results for each administration to provide an inter-nation as well as an intra-nation comparison. Secondly, the emphasis given to the use of the discrimination index in a certifying examination could be challenged. It could be expected that a large number of items should have low discrimination indices because the test is developed in a competency/mastery framework rather than one of differentiating candidates. Since the examinations of the Royal College are not at the stage of testing for competency, the decision was made to report the discrimination index.

Conclusions

The performance of the Australasian and Canadian candidates is very similar on the 160 item common examinations, even when the items are generated by a national rather than an international test committee.

Surprisingly, the items displayed stable cross-national characteristics and comparable performance regarding difficulty and discrimination. Differences in difficulty and discrimination do exist for some items, and these differences appear to be item specific, rather than nation-specific. In addition, differences might be due to lack of familiarity with the multiple choice question format. However, this does not appear to be the case, considering the very similar performances in the examinations.

Because of the confirmation of the relevance of the examination by the Australasian and Canadian content experts, it could be concluded that the test items are suitable for testing the knowledge of both Australasian and Canadian candidates seeking certification in Psychiatry.

Implications

From the results of this study and the one conducted by Burri et al., it would appear that there is merit in a program of international sharing of test material. While the need for local expertise and a national product cannot be denied, the duplication of effort by national medical testing institutions can hardly be justified.

References

Burri, A.T., Schumacher, C.F., and Vorkauf, H. The feasibility of using an American National Board examination for the evaluation of Swiss candidates for licensure. Medical Education, 11, 276-284, 1977.

39

PRECIS

PREDICTING CAREER OUTCOMES

MODERATOR: Marcia Z. Wile, Ph.D.
Case Western Reserve University, School of Medicine

THE ROLE OF PERSONAL THEMES IN MEDICAL SPECIALTY CHOICE

This paper argues that generalizations about the determinants of medical specialty choice need to be supplemented by a consideration of the student as an individual with particular concerns and issues. Case examples and content analysis of semi-structured interviews with senior medical students demonstrate that for each student personal themes can be delineated which influence the decision making process and shape student perceptions of medical school experience, role models, and various types of professional activity.

PREDICTION OF MEDICAL STUDENT CAREER CHOICE USING A FRESHMAN PERSONALITY PROFILE

A personality inventory administered to a Freshman medical school class, and the students' eventual specialty choice were subjected to a discriminant analysis. Results of the analysis indicate that this personality inventory may have predictive value for future medical school students' specialty choices, and is, therefore, a potentially valuable counseling tool.

PROXIMATE AND LONG TERM EFFECTS OF EARLY EXPOSURE TO PRIMARY CARE

Entering freshmen and sophomore medical students were trained to work as physician assistants in an HMO. This early exposure to primary care, HMOs and physician extenders resulted in no significant differences between the student PAs and non-PAs proximate measures of knowledge and attitudes, but did result in significant intermediate and long-term consequences, including subsequent electives, residency choice, and preferred practice arrangements.

Continued. . . .

THE IMPACT OF RESIDENCY ON PHYSICIAN PRACTICE PATTERNS: AN EXPLORATORY ANALYSIS OF YOUNG INTERNISTS

The relationship of residency training to the characteristics of practice of internists is investigated. The impact of training is chiefly imparted not by having participated in specific training procedures; rather, training in particular program environments such as large, research intensive programs (in contrast to small, clinically oriented programs) predicts the type of practice arrangements entered into and the allocation of time among various professional activities.

PHYSICIAN CAREER SATISFACTION: ANOTHER LOOK

This paper reports the results of a satisfaction with career test from a contemporary career study with 180 of the 1956-65 graduates of Case Western Reserve University School of Medicine. The career satisfaction scores are then compared with those from the comparable study of a decade ago with 1935-45 graduates, and some reasons are put forward for the changes in rankings on career satisfaction noted in several specialties.

The Role of Personal Themes in Medical Specialty Choice [1]

Tod S. Sloan and Sandra F. Bermann
Center for Research on Learning and Teaching
The University of Michigan

Medical students enter training with varying certainty as to the nature of the medical career they wish to pursue (Held and Zimet, 1975). Chief among the decisions they face is the type of medicine they plan to practice. Obviously, one's choice of medical specialty determines the basic structure of daily professional activity. On these grounds alone, medical specialty choice merits the attention of researchers in medical education.

The career decisions of medical students have gradually attracted the interest of investigators as the percentage of specialists increased drastically over the last fifty years (Kendall, 1971) and as the shortage of generalists came to be felt by the public (Funkenstein, 1979). Unfortunately, research on the determinants of medical specialty choice has been hindered by circumscribed methodological conceptions on the part of investigators in this area. In the interest of simplification and generalization important aspects of the person have been neglected in attempts to understand specialty choice. From the perspective of the deciding individual, the choice may involve considerations of status, finance, economic trends, availability of positions, role models, and life style as well as interpersonal, experiential, emotional, ethical, and intellectual factors. While any approach which taps into one of these realms may enlighten us to some extent about some students, it is difficult to bring such insight to bear when attempting to understand the choice of an individual. We will argue that a person-centered approach is essential if we hope to move beyond the generalizations provided by aggregate data.

Before introducing our data and method, we should refer to a few key studies which have shaped our thinking. Kendall (with Selvin, 1957; 1971) pioneered the research on specialization, showing that role models play an important part in specialty choice and that students move toward specialties in order to reduce the complexity and quantity of the material to be mastered. Zimet and his colleagues (Zimet and Held, 1975; Held and Zimet, 1975; McGrath and Zimet, 1977) have pursued a social psychological approach to understanding specialty choice. They have focused on aspects of the specialty such as prestige, money, intellectual breadth, social attractiveness, type of relation to patient, length of residency, and perceived similarity of self to typical specialists. These factors are important but not necessarily considered in the individual student's decision making process.

Funkenstein (1979) approached the topic from a broader sociohistorical perspective. His longitudinal study tied medical specialty choices to ideological and economic trends. He demonstrated that while we see these choices as originating in the individual student, they definitely reflect the operation of larger social trends. Funkenstein's analysis suggests that a matching process occurs between personal orientation to medicine, e.g., scientific, interpersonal, social service, and movements in the profession and in society. It would be useful to know more about this process for to hear it from the student's point of view such sociological factors are inoperative. We would suppose that these factors have a subtle, indirect impact on specialty choice by linking up with characterological aspects and personal ideologies of medical students as they are socialized into the profession. Our work in this paper is a step toward a conceptual orientation which will permit the investigation of these mediations between individual and society.

Method

The data from which we will draw to establish our point were gathered in order to compare the development of standard medical students at the University of Michigan with that of students in the experimental six-year AB-MD program (Inteflex). We conducted semi-structured interviews lasting approximately one hour with a sample of 90 members (68 males, 22 females) of a graduating medical school class. The interviews took place six months prior to graduation. By this time, most students had made the crucial decision regarding the type of internship they wished to pursue. The questions in the interview elicited responses concerning the strengths and weakensses of the medical school program, the effects of the medical school years on various aspects of the student's life and character, the student's plans for the future, expected satisfactions in the career, and current views of the profession.

In our first reading of the interview protocols, in preparation for coding and content analysis, we noticed that responses varied tremendously across students and that the interview could be viewed as a projective test: despite being enrolled in the same objective program, each student definitely brought his or her own concerns, conflicts, and values into that setting and reported these both directly and indirectly in the interviews. With this in mind, two independent coders extracted five or six personal themes for each interview. A personal theme was recognized when an issue or concern was expressed repeatedly across diverse interview topics. Intercoder reliability reveals a 90% agreement on at least three themes per student.

Case Examples

We will briefly describe three cases which demonstrate our approach.

Our first student, Robert, came to medical school expecting to be able to relax, to learn at his own pace and to follow his curiosity. In fact, he finds medical school to be just the opposite. He has to learn many things that are uninteresting and irrelevant to him. Robert became agitated by the demands of the school and dismayed with the lack of progressive educational techniques better suited to his independent style. In all, he feels that the school seriously hinders his personal development.

Thus far we know that this is a man who resents external constraints on freedom and values individuality, both his own and that of others. Robert attributes his not feeling well suited to this medical school to being a "free thinker" who should not have to prove himself to authority figures. He adds that the Inteflex program is turning out students who are unifaceted, impersonal and socially deprived. He claims to have a latent "Type A" personality within him and could, we think, if he were more tolerant of constraints, easily become quite compulsive. He is fighting hard to do well academically and to maintain his individual integrity at the same time.

Robert limits his comments on the practice of medicine to these complaints: he resents being told to practice medicine in particular ways, and dislikes hearing that a family practice career would be a waste of his potential.

Constraints on his sense of freedom is, once again, a thematic complaint. People who negatively stereotype the medical profession also bother Robert. He acknowledges that others' opinions affect him too much and contribute to his uncertainty about which medical field to enter. Curiously, he envisions financial security as one of medicine's most satisfactory aspects.

The two major themes of individuality and resentment of constraints hinder Robert in his attempts to commit to general medicine. Instead he opts for peripheral specialties such as ENT, ophthamology, and dermatology in order to "maintain outside interests" and to be near his family. He may be choosing a specialty of this sort to demonstrate his uniqueness or individuality and to avoid commitment to a field which is general or central to the medical world. The personal themes expressed in Robert's interview are called forth as a means of adapting, coping, surviving, and making the best of the medical school years. They are also salient formulators of lifetime career choice and hence determine the kind of job, income, lifestyle, and family experience that Robert will enjoy. More than representing economic and status factors, Robert's choice echoes both medical students' and society's current preoccupation with self-gratification (Lasch, 1979).

Each entering student brings to the medical school experience a stable set of personal themes and conflicts, even though the professional identity is still in flux (Held and Zimet, 1975). In Robert's case we can safely infer a characteristic concern for doing what he wants, resistance to authority and defense of individuality. The expression of these personal themes in the specialty choice seems to be a prerequisite to self fulfillment within that elected career. The life-structure toward which the student moves is bound to be a major means of satisfying the various and sometimes opposing aspects of the self.

In contrast to Robert, Bill expected to be taught in medical school. He found instead that he has to teach himself. He expected to be dependent but has had to develop independence and self motivation. Bill appreciates the school's prestigious faculty and dislikes the lack of direct teaching: "No one ever pulls the material together...there's never enough general information, only specifics...". While Bill feels well suited to this school he would trade some of the prestige for more direct help and feedback. His difficulty maintaining interest in tougher courses may indicate that Bill enjoys easy and immediately reinforcing activities.

He describes himself as obsessive-compulsive, competitive and sociable with interest in running track and mountain climbing. However, he feels that medical school has hindered these activities and rendered him more socially inhibited. Bill admires two medical role models, one for his display of motivation in climbing Mt. Everest and the other for her intelligence. As he approaches the last semester of school, he feels that his knowledge of the basic sciences is limited, that he will need support and guidance for some time to come.

We see that Bill needs a supportive atmosphere, both socially and intellectually and that he would prefer not to work very hard. His future as an orthopedic, sports physician in private practice may ensure that he will be surrounded by other athletes, thus reducing his social inhibition. He desires to practice in rural areas, or away from the scrutiny of more rigorous, intellectual physicians. The greatest gains of this career are envisioned by

Bill as short term treatment (i.e., no intimacy), correcting problems without having to think and attaining a slow and easy pace of life. When comparing Bill's resolution of thematic dilemmas with Robert's we see that Bill's solution is more satisfactory, leading to greater optimism and certainty.

A brief description of the third student reveals several salient themes or issues which, once again, are congruent to the choice of medical specialty. Susan admits that before entering medical school she had no concrete image of what doctors did but generally expected to learn in great detail. Her concern with learning is echoed in her list of the program's strengths and weaknesses. Unlike most peers, Susan is not pleased with the transition from basic science to clinical emphasis in the last two years and wanted more extensive coursework. Learning was evidently important to Susan before she came to medical school and continues to be a primary value.

The second general theme is Susan's strong scientific and weak interpersonal orientation. Her academic praise for peers, role models and program aspects far outweighs any mention of interpersonal qualities or patient interest. Self-described as selfish, with no close personal relationships, Susan's intellectual interests dominate her personality. A strain of passivity and helplessness is also evident throughout the interview data.

What career could Susan select which incorporates learning and science but does not require the skills of interpersonal interaction. Susan has chosen pathology because she likes the topic and is "particularly interested in diseases and tumors". Her future plans include an academic or teaching component which directly relates to the learning theme but may cause some friction with its interpersonal, student-teacher requirements. When asked to expound on the virtues of a teaching appointment Susan's first response is that students "force you to be active...they make you learn" material.

The advantages of a teaching and practice career in pathology coalesce Susan's personality themes and therefore serve her needs. She may find difficulty relating to students on an interpersonal level but will derive a mutual exchange on an intellectual level; i.e. the student stimulates, Susan teaches. In pathology, there is no patient contact and the emphasis is on disease and scientific problem solving.

These three students, although representing minor specialties, are representative in that each has brought to medical school a set of characterological concerns which pervade the interpretation of medical school experience and shape career interests. It is apparent that the nature of medical education is especially determinative in bringing certain issues into salience. For example, the time constraints of medical school warn some students to enter specialties which are less temporally constricting (see Edwards and Zimet, 1976).

Our emphasis on individuality in specialty choice may seem exaggerated to those who are aware of specialist stereotypes, e.g., that one should enter surgery in order to do something concrete and internal medicine for a greater intellectual challenge.

We tested the notion that students who chose the same specialty would be more alike than different in their interview issues or themes. Family or

General Practice was the branch selected by approximately one fourth of our sample. Many of these students share a humanistic or patient orientation and half describe a strong interest in acquiring knowledge or continued learning as important to them. But there are at least 45 other themes mentioned by only one or two students. Clearly this population is more differentiated than similar in their personal themes.

Again, for those in our sample who will enter surgery, there is more disparity than commonality among themes. Of the surgeons in our group, half share an interest in practical problem solving while such issues as ethical concerns, being efficient or striving for power emerge as unique or individual. While it is tempting to look at the higher percentage of a few common traits, exploration of that rich diversity and complexity of individual interest may, in the end, yield a greater understanding of the union of person and career.

Conclusions

To draw out fully the implications of our approach would require more space than we have. Instead, we will restate our primary points and encourage the adoption of a modified perspective in subsequent research on medical sepcialty.

The students whose decisions we studied here have struggled and, to varying extents, have resolved important personal issues in their specialty choices. These students are not atypical. Having to rule out large areas of medicine in order to specialize requires the student to confront his or her personal limitations. (From this perspective even general practice and family practice are specialization.) In the act of choosing a specialty students project themselves into an imaginary future and hope that it will be one that is satisfying and useful. The basis of that projection is the individual's unique organization of personal themes as magnified by the medical school experience. Any such choice in life is bound to incorporate private issues as much as it takes into account objective factors.

Medical students are fortunate in comparison to other young adults in that they have the luxury of choice, to mold a life-structure which provides for the satisfaction of personal needs while permitting a social identity which is highly esteemed, service oriented and financially rewarding. It seems to us that a greater explicit awareness of each individual in research on specialty choice may contribute to the improvement of decision making at the institutional and personal levels such that the private needs of physicians will support rather than interfere with the delivery of health care.

Our concern has not been to deny the obvious determinative power of demographic, economic, and role modelling factors in specialty choice. These ultimately shape the objective world in which the individual must negotiate a satisfying career. Moreover, these are the primary aspects of the medical profession to which policymakers can address themselves. In counseling or advisement settings, however, the individual dynamics which serve to mediate between social factors and individual choice should be considered with a view to improving the self-awareness of medical students as they make decisions which definitely affect their potential for career fulfillment in medicine.

----

[1] We acknowledge the contribution of Dr. Donald R. Brown, Director, Inteflex Premedical-Medical Program, to this paper in sharing his data and commenting on our analysis.

References

Edwards, M. T. and Zimet, C. N. Problems and concerns among medical students. Journal of Medical Education, 1976, 51, 619-625.

Funkenstein, D. H. Factors affecting career choices of medical students, 1958-1976. In E. Shapiro and L. Lowenstein (Eds.), Becomong a Physician. Cambridge, Mass: Balinger, 1979.

Held, M. L. and Zimet, C. N. Longitudinal study of medical specialty choice and certainty level. Journal of Medical Education, 1975, 50, 1044-1051.

Kendall, P. L. Medical specialization: Trends and contributing factors. In R. H. Coombs and C. E. Vincent (Eds.), Psychosocial aspects of medical training. Springfield, Ill,: Charles C. Thomas, 1971.

Kendall, P. L. and Selvin, H. C. Tendencies toward specialization in medical training. In R. K. Merton et al. (Eds.), The student-physician. Cambridge, Mass.: Harvard University Press, 1957.

Lasch, C. Culture of narcissism. New York, N.Y.: Norton, 1979.

Zimet, C. N. and Held, M. L. The development of views of specialties during four years of medical school. Journal of Medical Education, 1975, 50, 157-166.

47

PREDICTION OF MEDICAL STUDENT CAREER CHOICE FROM A FRESHMAN PERSONALITY PROFILE

Murray M. Kappelman, M.D.
Bernice Sigman, M.D.
Leslie Walker-Bartnick, M.A.

Purpose:

The purpose of this study was to predict senior medical students' actual career choices from a California Psychological Inventory taken at the beginning of their Freshman year. Defining a personality type likely to choose a specific specialty has important consequences to the counseling of medical students. The choice of a specialty is one of the most difficult and anxiety-producing decisions faced by medical students. If medical students have the available information concerning the specialty choices of previous medical students whose tests revealed similar personality characteristics, the student in medical school will have the opportunity to explore systematically these disciplines during his or her own medical school training.

The difficult career choice decision is generally necessary by the end of the third year of medical school so that the student can apply and interview for appropriate residency positions. For students in three-year programs, this process is even shorter. Because this decision affects the long-term goals of the student and often needs to be made before the student has the opportunity to become familiar with each specialty, many students would benefit from career counseling during the planning process of rotations and electives within their medical school training. Zimny and Senturia (1973) surveyed medical schools in a study of career counseling services and found a lack of formal services to aid students in selecting a medical specialty. Respondants to their survey indicated that in situations where counseling is done by faculty, in many instances students are recruited or given biased advice. One school mentioned that they have no system for advising advisors. These survey responses indicate two needs in medical education: (1) career counseling services for students, and (2) data-based information to advise the advisors.

In the 1960's, a flood of research on career choice by medical students appeared in the literature. Gough (1975) reviewed studies suggesting that preferences for different specialties may result from personality and motivational factors. Wunderlich and Gjerde (1978) and Plovnick (1979) studied the association between medical student learning style and their career choice. Gough and Hall (1977) looked at the career choice of medical students from non-medical families, and McGrath and Zimet (1977) found differences between the career choices of male and female medical students. While this research resulted in interesting findings, it did not provide a formula for predicting career preferences, and was generally based on data obtained at the end of medical school training when decisions were formulated or, in some cases, already made.

Send all reprint requests to Leslie Walker-Bartnick, M.A., Office of Medical Education, 10 S. Pine Street, Baltimore, Maryland, 21201.

One of the more promising variables which may predict interest in medical specialties is personality. Tucker and Strong (1962) designed an interest inventory with little success, but Gough (1975) in his review of the literature called for a "continuing flow of new information in order to supplement what prior studies have provided". Plovnick (1979) suggested that students are socialized into a career choice, and this premise implies that the use of psychological assessment may provide a means for understanding career choice patterns in medical school.

A well studied and much used psychological inventory is the California Psychological Inventory (CPI) developed in 1957 by Harrison Gough. It has been used in studying differences between male and female medical students (Cartwright, 1972) in predicting medical student performance (Gough & Hall, 1964; Korman, Stubblefield & Martin, 1968); stress in first year medical students (Boyle and Coombs, 1971); to name a few studies utilizing the instrument.

This research involves the use of psychological testing completed during freshman year to formulate a prediction of students' eventual career choice. The data, therefore, would be available early in the students' medical school training, and would provide an increased degree of objective information to the student and his advisor.

Methodology:

One hundred and seventy-five students entering the University of Maryland School of Medicine in the fall of 1976 were administered the California Psychological Inventory (CPI) on Orientation Day. The scales were machine-scored by Interpretive Scoring Systems.

The 480 item CPI was chosen because of its applicability to measuring personality characteristics important for social interaction in "normal" subjects. It has been the subject of reliability and validity research as well as personality studies. It also is easily administered to large groups. Profiles include standard scores on 18 scales; Dominance (DO), Capacity for Status (CS), Sociability (SY), Social Presence (SP), Self-acceptance (SA), Sense of Well-Being (WB), Responsibility (RE), Socialization (SO), Self-Control (SC), Tolerance (TO), Good Impression (GI), Communality (CM), Achievement via Conformance (AC), Achievement via Independence (AI), Intellectual Efficiency (IE), Psychological-mindedness (PY), Flexibility (FX), and Femininity (FE).

At the end of the senior year, in spring 1980, a discriminant analysis was performed on the students' selection of a residency and their scores on the 18 CPI scales. Because of attrition and leaves of absence, only 159 of the original 175 were 1980 graduates. Students fell into one of six groups; Primary Care (Medicine, Family Medicine, or Pediatrics), Surgery, Ob/Gyn, Psychiatry, Flexible, or a category designated as non-patient-care oriented (Radiology, Pathology, Pharmacology, Research, etc.).

The discriminant analysis provides classification coefficients for each CPI scale. Each classification coefficient is multiplied by the students' raw CPI score for that scale. The sum of these figures is added to a constant, resulting in a classification score for each specialty. The specialty group with the highest score for each individual case will be the highest probability specialty group for that student.

Results:

Prediction results of the analysis reveal that 48.3% of the students were correctly classified into their residency choice (highest probability group) using their CPI scores. However, in looking at students correctly classified by either their highest or second highest probability group, the percentage is 77% (see Table 1). Therefore, 77% of the class is correctly predicted within two groups.

The classification equations based on the data in this study are as follows:

Classification score (Medicine, Family Medicine, Pediatrics) = .07087DO+.21659CS +.08968SY+.45384SP+.55734SA-.25001WB-.08789RE+.56853SO+.90242SC-.54457TO+.20517GI +.98746CM-.45103AC+.18864AI+.06941IE+.33421PY+.55557FX+.60929FE-122.23849

Classification score (Surgery) = .12497DO+.20838CS+.03863SY+.52299SP+.60727SA-.22652WB-.18025RE+.52518SO+.94937SC-63348TO+.25424GI+1.03638CM-.49752AC+.18783AI +.10947IE+.35065PY+.58524FX+65342FE-131.14824

Classification score (OB/GYN) = .10582DO+.26337CS+.14181SY+.53718SP+.48946SA -.21117WB-.20333RE+.49202SO+1.07726SC-.64788TO+.11206GI+1.07081CM-.49258AC+.34915AI -.01169IE+.29725PY+52499FX+.63879FE-126.96041

Classification score (Psychiatry) = .12876DO+.36679CS+.01864SY+.39758SP+.66937SA -.36507WB+.00140RE+.56005SO+.91821SC-.60947TO+.15990GI+.97526CM-.44332AC+.02649AI +.19216IE+.36892PY+.71054FX+.60664FE-137.13568

Classification score (flexible) = -.01152DO+.38214CS+.23886SY+.24930SP+.49564SA -.26014WB-.29455RE+.59643SO+1.06729SC-.55749TO+.05277GI+.98669CM-.57360AC+.23472AI +.04043IE+.49029PY+.44901FX+.65756FE-113.96138

Classification score (non patient-care oriented) = -.06876DO+.34001CS+.16032SY +.33294SP+.58166SA-.17662WB-.03701RE+.53804SO+.87947SC-.58858TO+.19326GI+.90496CM -.42286AC+.17479AI+.01608IE+.39507PY+.48711FX+.51150FE-109.15231 where DO, CS, SY, SP, SA, WB, RE, SO, SC, TO, GI, CM, AC, AI, IE, PY, FX, FE are the students' raw CPI scale scores. These equasions can be used to predict the choice of all students who have taken the CPI regardless of their year in medical school.

Table I

PREDICTION RESULTS

Actual Group	N of Cases	Number & Percent Correctly Predicted (Highest Probability)	Number & Percent Correctly Predicted (2nd Highest Probability)	Total
Medicine Pediatrics Family Medicine	99	38 (38.4%)	38 (38.4%)	76 (76.7%)
Surgery	24	13 (54.2%)	5 (20.8%)	18 (75%)
Ob/Gyn	23	14 (60.9%)	3 (13%)	17 (74%)
Psychiatry	4	4 (100%)	---	4 (100%)
Flexible	4	4 (100%)	---	4 (100%)
Non-patient	5	4 (80%)	---	4 (80%)
Total	159	77 (48.43%)	46 (28.93%)	123 (77.36%)

Discussion & Implications:

The results of this study indicate that a CPI taken as early as Orientation Day of the Freshman year of medical school can be predictive of residency choice for 77% of the students in this study. Using this type of data, counselors could inform students of their two highest probability career choices during Freshman year. Undecided students could explore career choices with an emphasis on their two highest probabilities by selecting related elective experiences during their medical school training.

While the sample sizes choosing Psychiatry, Flexible, and non-patient-care oriented residencies are small, their prediction accuracy is higher than the other specialties, and their selection by students is generally much below that of Surgery, Medicine, Ob/Gyn, Pediatrics, and Family Medicine. It will be interesting to add further variables already proven to have predictive value for career choice to the CPI data in an effort to improve the predictability rate of 77%.

51

References

Boyle, B.P. & Coombs, R.H., Personality Profiles Related to Emotional Stress in the Initial Year of Medical Training, Journal of Medical Education, 1971, 46, 882-888.

Cartwright, L.K., Personality Differences in Male and Female Medical Students, Psychiatry in Medicine, 1972, 3, 213-218.

Gough, H., Manual for the California Psychological Inventory, Consulting Psychologists Press, Palo Alto, California, 1957.

Gough, H. & Hall, W., Prediction of Performance in Medical School from the California Psychological Inventory, Journal of Applied Psychology, 1964, 43, 218-226.

Gough, H., Specialty Preferences of Physicians and Medical Students, Journal of Medical Education, 1975, 50, 581-588.

Gough, H., and Hall, W., A Comparison of Medical Students from Medical and Non-medical Families, Journal of Medical Education, 1977, 52, 541-547.

Korman, M., Stubblefield, R., & Martin, L., Patterns of Success in Medical School and their Correlates, Journal of Medical Education, 1968, 43, 405-411.

McGrath, E., and Zimet, C., Female and Male Medical Students: Differences in Specialty Choice Selection and Personality, Journal of Medical Education, 1977, 52, 293-300.

Plovnick, M., A Longitudinal View of Learning Styles and Career Choices, Journal of Medical Education, 1979, 54, 245-246.

Tucker, A. and Strong, E., Ten-Year Follow-up of Vocational Interest Scores of 1950 Medical College Seniors, Journal of Applied Psychology, 1962, 46, 81-86.

Wunderlick, R. & Gjerde, C., Another Look at Learning Style Inventory and Medical Career Choice, Journal of Medical Education, 1978, 53, 45-54.

Zimny, G., and Senturia, A Medical Specialty Counseling: A Survey, Journal of Medical Education, 1973, 48, 336-342.

PROXIMATE AND LONG-TERM EFFECTS OF EARLY EXPOSURE TO PRIMARY CARE

Marian Osterweis, Ph.D. and David L. Rabin, M.D., M.P.H.
Georgetown University School of Medicine

1. Purposes and Hypotheses: Entering freshman and sophomore medical students were trained during an 8-week summer course to be physician assistants in an HMO. Elective time during the next year was spent working as PAs in the HMO in which they were trained. The purposes of the program were: (1) to give students early clinical exposure to primary care, HMOs, team practice, and the roles of physician extenders; and (2) to teach them to communicate effectively with patients and providers. Short-term, intermediate and long-term effects of the program were hypothesized:

(1) In the short run, the program would result in increased knowledge about and more positive attitudes toward primary care, HMOs, and physician extenders.
(2) The program would have intermediate benefits for students in terms of clinical and communication skills which would be apparent in their third and fourth years.
(3) In the long run, the program would affect career decisions including specialty choice and practice arrangements.

2. Background and Critical Review: Social reforms in the 1960s led to major federal initiatives to encourage more equality in medical care and more education in primary care. Many commentators had found the traditional system lacking--medical students and graduate physicians were insensitive to human needs (1), medical practice was not relevant to societal needs (2), university medical centers were unresponsive to their surrounding communities, and medical students were trained in isolation from other health professionals with whom they were later expected to work cooperatively (3). In response to these criticisms, 30 new medical schools were established in the U.S. and Canada with the explicit intention of training medical students in the community and for the community. Traditional schools responded initially at the graduate level with residency programs in Family Practice and other primary care specialty programs. Subsequently, undergraduate medical school curricla were altered to provide primary care training. Recent federal manpower support has been conditional upon medical school success in influencing decisions towards primary care. It is difficult to obtain time in the undergraduate curriculum for general training as medical science continues to expand and as further subspecialization continues to occur. However, since at least 50% of future graduates should enter primary care in order to care for our population, and since attainment of this goal may be legislatively mandated, it is ever more critical that we not only make curriculum changes but that these changes effect the desired results in terms of behaviors.

Medical practice is in a state of flux with larger numbers of physicians being salaried, continued growth of both prepaid and nonprepaid group practice, and increased numbers of trained physician extenders. The medical graduates are entering into a rapidly changing and organizationally ever more complex environment. It is therefore appropriate for medical schools to alter their curricula in order to work effectively in this environment (4).

Reprint requests to Dr. Osterweis, Dept. of Community & Family Medicine, Georgetown University School of Medicine, Washington, D.C. 20007.

Many studies have been undertaken to assess changes in medical students' attitudes as they progress through medical school (5,6), to determine the factors which affect career choice (7,8), and more recently to develop programs whose purpose is to affect attitudes and career choice. The latter usually have as their goals the production of more humanitarian, socially conscious, primary care practitioners. Although it is well accepted that attitudes do change in the course of medical training, there is some disagreement as to whether those changes are towards more cynicism or more humanitarianism. Career preferences also fluctuate, but attempts to relate these to any set of background factors such as hometown size, parents' education, religion, college major, etc., have produced contradictory findings. It is also not clear at what stage students are most impressionable so that programs aimed at influencing attitudes and practice choice exist at all levels of medical school training. Finally, there is little in the literature which discusses the reasons why students choose to participate in these programs in the first place and what the long-range effect of them is because evaluations tend to be conducted soon after the exposures, if at all.

3. Methods: This was a five-year program in which equal numbers of students from the entering freshman and sophomore classes were randomly chosen from among those who volunteered. The program grew from 4 students from each class the first year, to 8 the second, 12 for the third and fourth years, and 11 for the final year, making a total of 94 program participants.

The 8-week summer program was based on the Automated Military Outpatient System (AMOS) of algorithms which codify care of common disorders. During the initial three weeks of intensive didactic training, students were instructed in the use of algorithms; they received training and practice in physical diagnosis and communication skills; and they attended lectures on the diagnosis and management of common primary care problems, HMOs, and the roles of nonphysician providers in primary care. This was followed by five weeks of clinical work in the HMO, initially with physician preceptors, and in the last two years with midlevel practitioners who were either physician assistants or nurse practitioners. As the program evolved, increased emphasis was placed on communication skills, patient education and personal identification with physician extenders.

The evaluation was designed to assess initial differences due to self-selection bias, proximate, intermediate and long-term program effects, using a three-group pretest/repeated posttest design. In each year, substantially more students applied for the program than could be accepted. Since some initial differences based on self-selection factors were hypothesized, those who were not randomly selected for the PA program were kept for evaluation purposes as one of the control groups--hereafter called "volunteers". An equal number of "controls" were randomly selected from among those students who did not wish to be in the program. All PAs, volunteers and controls were given a pretest in the summer or early fall to discern background differences, knowledge of and attitudes toward primary care, HMOs, and physician extenders. The same questionnaire was used as a posttest at the end of the sophomore year. Finally, PAs and controls were questioned in their fourth year about residency choice, preferred practice arrangements, and primary care electives during their clinical years. PAs were also asked about program effects on specific clinical areas at that time after having completed the bulk of their clinical training. Response rates varied from 100% for the PA pretests to 65% for the control group first posttests.

Data were analyzed using Chi square to determine whether there were differences between PAs and the other groups, and to see whether program impact varied by preceptor type or by timing of the program.

4. Results: Based on what was known from the literature, it was hypothesized that:

(1) Volunteering for the PA program would be related to three background variables: undergraduate major, parents' education, and whether one was raised in an urban or rural setting.
(2) Those who volunteered for the program would have greater knowledge about and more positive attitudes toward primary care than those who did not volunteer.
(3) Sophomores would know more about primary care than freshmen.
(4) Freshmen attitudes would be more positive than sophomores toward primary care.
(5) Knowledge and attitudes would increase as a result of the program.

None of these five hypotheses was confirmed. Although there were some small differences in the predicted direction, none was significant. Background and class year have little effect on pretest knowledge or attitudes. The program appeared in the short run to have no effect on knowledge or attitudes, regardless of whether students participated in their freshman or sophomore year and regardless of whether their preceptors were physicians or midlevel practitioners.

In order to assess intermediate range effects, students were asked about areas in which the PA experience in their preclinical years was an advantage to them in subsequent clinical years. As can be seen in Table 1, the program was clearly useful to students in terms of communication skills, understanding broad patient needs, knowledge of other providers, and self-confidence. The perceived advantages of the program were greater in the third year than in the fourth year as one might expect.

Most striking, however, are the long-term follow-up data regarding the PA program's impact on career choices. Table 2 shows that the students from the PA program are more likely to go into primary care, to want to practice in HMOs, to want to be salaried, and have taken more electives in primary care. In addition, the PA experience is perceived as having been more influential in each of those decisions than other primary care experiences had by the controls. (Chi square tests of significance were run for each item comparing freshmen with sophomore PAs and comparing PAs as a group with controls. Those which reached statistical significance are marked with an asterisk.)

Additional analyses compared students who were precepted by physicians (years 1 to 3 of the program) with students who were precepted by physician extenders (years 4 and 5 of the program). No significant short, intermediate or long-term effects were found when controlling for preceptoring. (Not all of our final follow-up data is in for the final year of the program, so the conclusion about long-term effects is still tentative.)

5. Discussion: From these data, it would appear that early clinical exposure to primary care has profound but delayed effects. Many students commented in open-ended questions that although the program gave them a "leg up" for their third year especially, it was not until they had completed the bulk of their clinical rotations that they realized how influential the program had been. Most were initially interested in the program because it was a summer job and/or was an opportunity for early clinical experience. Some already knew they were interested in primary care. Only one had any intention of practicing in an HMO

setting initially. Many students (including those who were and were not going into primary care specialties) commented that the program helped them make realistic, educated career decisions.

6. Conclusions and Implications: This final two years of follow-up data confirms our previously tentative conclusions about the impact of early exposure to primary care on later career decisions. The primary care experiences offered in traditional medical schools such as ours are so limited and expose students to something so different from the rest of their clinical experiences that it takes time to digest the experiences and put them in perspective. Career choices reflect the total experience of medical school. Since the form and content of this program evolved over a 5-year period, but the effects were the same for each year, it would appear that the key factor is early exposure. If the opportunities for primary care experiences in group and team practice are limited, it would seem that the most effective time to have them is during the "preclinical" years in order to influence later attitudes and behaviors.

Despite the difficulties in following students over time, evaluators of these programs should plan longitudinal assessments of program impact. Assessments immediately following such programs typically show very little. Perhaps many of the programs cited in the literature as being ineffective catalysts were in fact more influential than we know in terms of ultimate behaviors and attitudes.

References

1. Smyth, F.S. The place of the humanities and social sciences in the education of physicians. J. Med. Educ. 37:495-499, 1962.

2. Hunt, A.D. A time of change and reform in education. In Medical Education Since 1960: Marching to a Different Drummer, A.D. Hunt and L.W. Weeks (eds.), Lansing, Mich. State Univ. Foundation, 2-9, 1979.

3. Baldwin, D.C. and Baldwin, M.A. Interdisciplinary education and health team training: A model for learning and service. In Medical Education Since 1960: Marching to a Different Drummer, A.D. Hunt and L.W. Weeks (eds.), Lansing, Mich. State Univ. Foundation, 175-189, 1979.

4. Freidin, R.B., Barsky, A.J., Levine, D., and Williams, S.R. Medical education and physician behavior: preparing physicians for new roles. J. Med. Educ. 47:163-168, 1972.

5. Rezler, A.G. Attitude changes during medical school: a review of the literature. J. Med. Educ. 49:1023-1030, 1974.

6. Merton, R.K., Reader, G.G. and Kendall, P.L. (eds.). The Student Physician. Cambridge, Harvard Univ. Press, 1957.

7. Gough, H.G. and Ducker, D.G. Social class in relation to medical school performance and choice of specialty. J. Psychol. 96:31-43, 1977.

8. Herman, M.W. and Veloski, J. Family medicine and primary care: trends and student characteristics. J. Med. Educ. 53:99-106, 1978.

TABLE 1

Area in Which PA Experience Was an Advantage During Subsequent Years of Training

	THIRD YEAR		FOURTH YEAR	
	Fresh. PA	Soph. PA	Fresh. PA	Soph. PA
Specific clinical skills	85%	80%	33%	44%
Awareness of patient behavior	89%	86%	50%	67%
Knowledge about common and minor disorders	81%	91%	50%	89%
Communication skills with physicians	63%	63%	42%	39%
Communication skills with other providers	74%	63%	50%	50%
Communication skills with patients	89%	89%	58%	50%
Understanding of needs of chronic patients	26%	35%	0	28%
Self-confidence	70%	83%	58%	39%
Knowledge about physician assistant roles	81%	83%	67%	78%
Knowledge about nurses' roles	44%	54%	17%	33%
Knowledge about other health professionals' roles	70%	46%	58%	39%
Awareness of social/emotional needs of patients	59%	63%	50%	39%
Sensitivity to family interactions	59%	51%	42%	17%
Awareness of the needs of patients following hospitalization	22%	14%	17%	17%
knowledge of community resources for patient referral	52%	31%	42%	28%
content of secondary and tertiary care	48%	40%	42%	11%
N =	27	35	12	18

TABLE 2

Summary of Final Follow-up Data by Program Status

		Fresh. PAs N = 27	Soph. PAs N = 35	Total PAs N = 62	Controls N = 64
Residency:	Family Practice	7 (26%)	4 (11%)	11 (18%)	6 (9%)
	General Medicine	1 (4%)	13 (37%)	14 (23%)	9 (14%)
	General Peds.	1 (4%)	0	1 (2%)	3 (5%)
	Total Primary Care	9 (33%)	17 (49%)	26 (42%)	18 (28%)
(for controls only - ambulatory experience)		n/a	n/a	n/a	33 (52%)
Residency influenced by PA or amb. experience		13 (48%)	20 (57%)	33 (53%)	13 (39%)
*Preferred Practice Setting:	Hospital	2 (7%)	8 (23%)	*10 (16%)	*14 (22%)
	Solo	0	4 (11%)	4 (6%)	8 (13%)
	Group	18 (67%)	19 (70%)	37 (60%)	37 (58%)
	HMO	7 (26%)	4 (11%)	11 (18%)	1 (2%)
				x^2=11.92, 2 d.f., p<.01	
*Setting influenced by PA or amb. experience		16 (59%)	22 (81%)	*38 (61%)	*8 (13%)
				x^2=11.84, 1 d.f., p<.001	
*Payment Preference: (omitting hospital based practice)	Salaried	15/24 (63%)	10/27 (37%)	*25/51 (49%)	11/51 (22%)
	Fee-for-service	9/24 (38%)	17/27 (63%)	26/51 (51%)	35/51 (69%)
				x^2= 6.53, 1 d.f., p<.01	
*Payment preference influenced by amb. exper.		18 (67%)	17 (49%)	*35 (56%)	*7 (21%)
				x^2=10.84, 1 d.f., p<.001	
*Senior electives in primary care		*28/89 (31%)	*16/91 (18%)	44/180 (24%)	45/209 (22%)
		x^2= 4.69, 1 d.f., p<.05			

The Impact of Residency on Physician Practice Patterns: An Exploratory Analysis of Young Internists

BY: Peter A. Weil, Ph.D.; Mary Kay Schleiter, M.A.; Alvin R. Tarlov, M.D. (Department of Medicine, University of Chicago) and Robert C. Mendenhall, M.S.; Christy Moynihan, Ph.D. (Division of Research in Medical Education, University of Southern California School of Medicine)

Purpose of the Research

It has often been suggested by educators and researchers alike that the characteristics of the physician's formal training and clinical experiences have a direct influence on the mode(s) of practice adopted after completion of training. Whether or not this is true is obviously important to those concerned with training and to those whose interests lie in planning for the optimal use of physician manpower. In fact, any effects of the training received upon practice patterns should be known as they would suggest factors in the delivery of medical services which are amenable to change -- assuming, of course, that change of some sort is desired.

Two central questions guided the inquiry:

A. Does having attended a particular type of residency training program predict the practice characteristics of physicians?

B. What specific training procedures which take place during residency correlate with subsequent practices of physicians? Three areas are studied:

   1. Does the relative emphasis of a particular subspecialty during residency affect the chief diagnoses encountered in practice?
   2. What is the impact of the following ambulatory training features on subsequent practices:
      (a) ambulatory training in a variety of specialties
      (b) constituting residents into "practice groups" to provide ambulatory care
      (c) devoting relatively more time to ambulatory training (and less time to inpatient and consultative activities)
   3. Does the amount of longitudinal patient care provided by the resident in training correlate with the amount of continuous, longitudinal care provided in subsequent practice?

Review of Literature

This research can be viewed as part of a long tradition of studies on the socialization of physicians. Specifically, there are two competing paradigms currently offered: (1) the functionalist or Mertonian tradition and (2) the structuralist (Becker) critique. The functionalist school maintains that it is the process of medical education which underlies practice behavior. The structuralist school argues the organization of the immediate practice environment (e.g., practice arrangement) governs physician behavior (1, 2). Our research is an exploratory test of the impact of residency on physicians' future practices.

Previous research relating residency training to subsequent practice has primarily focussed on career decision making, notably specialty pursuits and

---

Address reprint requests to: Peter A. Weil, Ph.D.; Box 273, University of Chicago Hospitals, 950 E. 59th Street; Chicago, Illinois 60637

commitment to academic medicine and clinical practice (3-5). A limited number of studies have attempted to relate graduate medical training to the quality of medical practice (6-9) and utilization of clinical and technical services (10-12). Those studies, however, suffer from very crude characterizations of the residency training programs. For example, Petersen's study of general practitioners in North Carolina measured training programs based on the closeness of the teaching hospital's affiliation with a medical school. The study revealed no relationship between the type of internship or residency program taken and the quality of practice (6). A more recent study which attempted to relate graduate training to use of technical and clinical services by practicing physicians also suffers from a very general characterization of residencies (12).

This research utilizes the analyses of a prior study of the universe (418) of residency training programs in internal medicine (5). The data collected permitted the creation of a typology in which all 418 programs were assigned to one of five types. The purpose of the typology was to systematically characterize the differing environments of internal medicine teaching programs in the U.S. Key differentiating dimensions which were utilized to categorize programs included aspects related to the program's size; its internal differentiation; affiliation with other hospitals and medical schools; types of role models made available; and specific activities of the residents. Large, research intensive programs were classed as Type I and small, clinically oriented programs were classed as Type V. Types II, III, and IV took on intermediate values (5). The survey of training programs probed specific procedures of resident training: in the subspecialties, in ambulatory care and in longitudinal care of patients.

Other studies attempting to relate a composite measure of training environment to practice patterns have not been reported. Moreover, the impact of specific training procedures on practice are likewise few. Only one recent study was discovered which suggested that training in specialties other than adult medicine was directly related to the quality of adult care provided (8). In contrast, the research reported here attempts to measure whether specific procedures of training carry over into practice, or whether specific practice patterns are more closely correlated with the structural forces of the physician's practice arrangement.

Assumptions and Methods

The residency program data were collected in 1976-1977 from all (n=418) approved training programs in internal medicine. Though 409 programs responded (98%), information obtained by telephone provided information on the remaining 9 programs. At the same time that the residency program survey was being conducted, a stratified random sample of 5983 internists was selected for study of practice patterns from the American Medical Association's Physician Masterfile. Of these, 3868 completed and returned the practice log diary for an overall response rate of 65%. Among the individual specialties, the response rates ranged from 53% for General Internists to 82% for Endocrinologists (13).

In addition to sampling by subspecialty, stratification was based on the physician's type of practice and involvement in patient care services. The first three strata (Solo, Partnership and Group) conform to conventional thinking regarding "office based" practices. The fourth stratum, institutional, is composed of physicians employed in hospitals and medical schools who are predominantly involved in direct patient care. The fifth stratum, other, is composed of physicians who are not primarily involved in direct patient care (e.g., teachers, researchers or administrators) or who provide direct patient care under some type of practice arrangement not in the first four strata.

Because both residency program data and practice characteristics were collected in the same year, this study rests on the assumption that characteristics of training programs are relatively stable and do not alter radically over time. To increase the likelihood of measuring the impact of residency programs' effect on practice, we confined the sample to practicing internists who had graduated from medical school no earlier than June, 1963. This would allow no more than ten years in practice subsequent to the completion of residency. In all, the sample amounted to 193 General Internists and 781 internists who practice in the ten recognized subspecialties of internal medicine.

Another assumption implied by this research is that the relationships which are found between training and practice are, in fact, due to training and not due to the demands of current practice arrangements. To account for the possible confounding effects of this intervening variable, special statistical techniques were employed. For example, in testing the effect of the residency program typology we attempted to control for the demands of specific practice arrangements by ordering our independent variables so that variation due to other than typology effects were removed prior to its estimation. Also, in correlational analysis, partial correlations were obtained for the same purpose.

Results

Major findings of the impact of type of residency training program on subsequent practice patterns are specified by those who classify themselves as General Internists and those practicing in one of the subspecialties of internal medicine. Ignoring the possible effects of pre-selection of individuals predisposed to certain career paths into the varied residency program environments, the following represent the significant findings:

A. Relationship of Residency Typology to Practice

Subspecialty internists from larger, highly differentiated and research intensive residency programs (Types I and II) tend to enter non-office based practice settings -- particularly academic settings. These effects, while significant for subspecialists, were found to be insignificant for general internists, using a Chi Square test of significance.

Type of training program was related to the allocation of time to various activities in practice. Subspecialty internists from Types I and II residency programs spent significantly more time in teaching, research and 'other administration' (i.e., purchasing, personnel and management) and less time in direct patient care when compared with subspecialists who had attended Types III, IV and V programs. These variations in activities held controlling for type of practice arrangement. These trends were apparent

~~but~~ not as strong for General Internists.

Subspecialists trained in smaller, less differentiated training programs saw more patients per week and more patients per professional hour than the internists from Types I and II programs. However, the differences disappeared when the time spent providing direct patient care was controlled. Thus productivity in provision of patient care above and beyond the differences in the amount of time devoted to patient care is not associated with the residency training program typology.

Subspecialty internists from Types III-V residency programs (smaller and less differentiated) tended to provide more "primary" care to their patients than those from Types I and II programs. These effects disappeared however, when type of practice arrangement was controlled.

B. Relationship of Specific Residency Training Program Features to Practice

SUBSPECIALTY EMPHASIS IN TRAINING AND IN PRACTICE

No correlation was observed between the relative amount of training taken in any subspecialty during the residency and the proportion of patients who were diagnosed with diseases in that subspecialty's diagnostic code. Hematology training was the exception to this general finding.

AMBULATORY CARE IN TRAINING AND IN PRACTICE

- Providing ambulatory training in non-internal medicine specialties was correlated with general internists': (1) providing the majority of care for patients, (2) spending more time in patient care activities, and (3) allocating more time to ambulatory care.
- Organizing residents into practice groups was correlated with general internists': (1) spending more time in the ambulatory care setting, (2) providing more care in the ambulatory care setting, (3) providing more specialized care, (4) responding to consultative requests, and (5) teaching activities
- The amount of time in ambulatory training for both general internists and subspecialists was unrelated to the amount of time spent in ambulatory practice. Controlling for the type of practice arrangement, a positive correlation existed for generalists in solo practice. (For all other internists in other practice arrangements, this relationship continued to remain insignificant.) The amount of time in ambulatory training for both general internists and subspecialists was <u>negatively</u> related to provision of primary care in practice.

LONGITUDINAL CARE IN TRAINING AND IN PRACTICE

The proportion of longitudinal care residents provide during their training is positively correlated for General Internists with: (1) the proportion of patients in the practice who were regular patients; (2) the provision of principal care (and the correlation is negative with first encounter, episodic, specialized and consultative care); and (3) for subspecialists, the correlations are inconsistent and with some measures opposite of those patterns displayed by General Internists. (For explicit definitions see Reference #14.)

The proportion of regular patients in the practice was examined with four residency program measures of longitudinal care controlling for type of practice arrangement. While some relationships remained, others were no longer significant. These results gave clear evidence that the type of practice arrangement acts to specify the impact of longitudinal training on practice. The residency training program environment appears to mediate the aforementioned correlates of longitudinal care by affecting the type of practice arrangement which subspecialists enter and which may affect general internists as well.

Summary and Conclusions

Overall, we found support for the differential impact of a residency program typology which had been previously developed by our research group. Its impact is chiefly seen in terms of the type of practice arrangement which subspecialists enter into. Those from large, differentiated programs more often select to practice in non-office based settings and in academia. Moreover, even controlling for the type of practice setting, the time devoted to various activities seems to vary systematically for persons from the various program types. This was true for subspecialists in internal medicine and also to some extent for General Internists. Both groups who had trained in large, complex residency programs devote more time to teaching, research and other administrative activities such as management, and less time to patient care than those from smaller, less differentiated and more clinically-oriented training programs.

These findings attest to the theoretical assertions of the functionalist school of medical education -- i.e., that norms, values and even preferred activities imbued during training carry over into the physician's professional careers. Not controlled in this study, however, are the probable effects of pre-selection factors affecting entry into the varied types of residency training programs. What this exploration study did affirm is that type of residency program attended successfully predicts some fairly global parameters relating to an internist's subsequent career.

Our attempt to relate more specific aspects of residency training to subsequent practice patterns was notably less successful. First, emphasizing one or another subspecialty in training did not correspond to the types of principal diagnoses discerned in practice. Second, those who had attended training programs devoting more time to ambulatory care revealed no relationship to time spent in ambulatory care practice at first, and even after controlling for the type of practice arrangement, the relationship was significant only for General Internists in solo practice. In fact, the strongest correlate of time devoted to ambulatory care in practice was having taken

Third, longitudinal care provided by trainees was related to the proportion of regular patients in the practice for General Internists but the correlations for subspecialists were inconsistent. Controlling for the type of practice arrangement, the previous associations discovered were <u>specified</u> i.e., they continued to be significant in certain practice arrangements but were not significant in others.

The results of this exploratory study allow us to conclude that specific experiences incorporated into residency programs may or may not have an impact on subsequent practice patterns. At least one key intervening variable seems to affect patterns of practice i.e., the type of practice arrangement. The structuralist school gains ground if these findings relating specific training activities to practice are to be believed. That is, the specific demands of the practice setting are apparently important proximate factors influencing actual behavior which act to specify elements contained in prior graduate training.

The policy import of all this can be suggested: to legislate specific requirements to be met by training programs with the hope of influencing future practice patterns will by itself provide a feeble lever for change. More potent potential for changing practices lies in nurturing one or another of the distinctive types of training program environments, possibly to be combined with other more specific training initiatives not encouraged in existing legislation.

References

1. Merton RK: "Some Preliminaries to a Sociology of Medical Education." In Merton G. Reader and P. Kendall (Eds.) The Student Physician. Cambridge: Harvard University Press, 1957, pp.77-78.
2. Becker HS, Geer B, Hughes EC and Strauss A: Boys in White. Chicago: Aldine, 1961, p.432.
3. Anderson R Bruce W: Choosing a medical specialty: a critique of the literature in light of 'curious findings'. Journal of Health and Social Behavior 16:152-162, 1975.
4. Erdman JB, Jones RF, Tonesk X: AAMC Longitudinal Study of Medical School Graduates of 1960. Washington, D.C., Association of American Medical Colleges, 1978.
5. Weil PA, Schleiter MK, and Tarlov AR: National study of internal medicine manpower: II. A typology of residency training programs in internal medicine. Ann Intern Med 89:702-715, 1978.
6. Peterson OL, Andrews LP, Spain RS, and Greenberg BG: An analytical study of North Carolina general practice. J Med Educ 31:pt.2, 1956.
7. Clute KF: The General Practitioner: A Study of Medical Education and Practice in Ontario and Nova Scotia. Toronto: Toronto University Press, 1963.
8. Morehead MA and Donaldson R: Quality of clinical management of disease in comprehensive neighborhood health centers. Medical Care 12:301-315, 1974.
9. Palmer RH and Reilly MC: Individual and institutional variables which may serve as indicators of quality of medical care. Medical Care 17: 693-717, 1979.
10. Schroeder SA, Kenders K, Cooper JK and Piemme TE: Use of laboratory tests and pharmaceuticals. JAMA 225:969-973, 1973.
11. Freeborn DK, Baer D, Greenlick MR, and Bailey JW: Determinants of medical care utilization: physicians use of laboratory services. Am J Pub Health 62:846-853, 1972.
12. Pineault R: The effect of medical training factors on physicians utilization behavior. Medical Care 15:51-67, 1977.
13. Girard RA, Mendenhall RC, Tarlov AR, Radecki SE, Abrahamson S: A national study of internal medicine and its specialties: I. An overview of the practice of internal medicine. Ann Intern Med 90:965-975, 1979.
14. Mendenhall RC, Tarlov AR, et.al.: A National Study of Internal Medicine and Its Specialties: II. Primary care internal medicine. Ann Intern Med 91:275-287, 1979.

PHYSICIAN CAREER SATISFACTION: ANOTHER LOOK

Betty Hosmer Mawardi, Ph.D.
Case Western Reserve University School of Medicine
Cleveland, Ohio

BACKGROUND

Satisfaction of physicians with their careers in medicine has attracted considerable attention from those interested in the education and development of doctors. This paper reports the preliminary results of a job satisfaction test used in a career study with medical graduates from the classes of 1956-65 at Case Western Reserve University and then compares these findings with those from a previous study conducted with similar subjects a decade ago.

In addition to job satisfaction, the career study, in its entirety, has investigated other career aspects such as styles of practice, methods of maintaining medical competence, and the graduates' evaluations of their innovative medical education program.(1) Satisfactions of the physicians with their careers in medicine was also a prominent feature in a previous career study of graduates from the classes of 1935-45 at Case Western Reserve.(2)

METHODOLOGY

Subjects in the present career study included 180 of the approximately 800 graduates from the first ten classes of the revised program in medical education. They were chosen by a stratified random procedure that selected 20 in each of 9 different categories--general or family practitioners, internists, general surgeons, psychiatrists, pediatricians, obstetrician-gynecologists, full-time medical school faculty, full-time hospital based personnel, and women physicians in different specialties. The total of 180 included 160 male physicians and 20 female physicians. The females are somewhat more over-represented than they were in the total population for those years when women comprised only 8% of the student body. The overall methodology encompassed a longitudinal format and these graduates were studied repeatedly when they were students; however, the present paper reports only their present job satisfaction with their medical careers.

Participation in this present study, conducted in the late 1970's and 1980, involved a personal interview in the physician's office, wherever that may have been, and the completion of a booklet of additional short answer materials plus several tests, including one of job satisfaction. At the time of the submission of this paper, 142 booklets, or 79%, have been returned. We hope to have more booklets returned (in the previous study, about 90% of the booklets were returned eventually) and the report will be updated in the fall to incorporate any additional booklets that come in. It was decided that a 79% return is sufficient data to report at this time.

In the earlier study, conducted in the late 1960's, there were 153 male graduates, approximately 20 in each of the first 8 categories named above. There was no sample of female graduates in that investigation because, at the same time, Glick was studying all of the women graduates of the medical school as a part of her work for a doctoral degree.(3)

Reprints: Betty Hosmer Mawardi, Ph.D., Division of Reseach in Medical Education, Case Western Reserve University School of Medicine, 2119 Abington Road, Cleveland, Ohio 44106

Both major career studies of the 1956-65 and 1935-45 graduates employed the same methodology throughout, and both studies assessed career satisfaction by means of a modified Brayfield Job Satisfaction Blank (4) wherein the range of possible scores lies between 18 and 90. The modification in the test was the use of the work "profession" instead of "job" since it was felt that the former term was more appropriate for physicians. This modification has been used in other investigations and Brayfield indicates that it is valid but tends to raise the individual scores several points.

RESULTS

Table 1 shows the mean scores for all groups in both studies with the 1956-65 and the 1935-45 graduates. Table 2 indicates the rankings of all groups in both studies and compares the changes in ranks.

In examining Table 1, one notices first the virtual sameness in the total averages for all of the 8 groups of male physicians, 75.03 for the 1956-65 groups and 75.12 for those of the 1935-45 years. Mean scores for the 1956-65 groups ranged from 79.08 to 73.07 whereas the mean scores for the 1935-45 groups went from 80.18 to 69.85. Sub-totals for the 6 groups of practitioners in each study were almost identical as a whole. The 2 groups termed "non-practitioners" (full-time medical school faculty and full-time hospital based personnel) had total means essentially the same also. The more interesting data, however, are gleaned from the shifts in rankings of the groups noted in Table 2.

Today, ranking first of all groups in job satisfaction, are the general surgeons. Previously, general surgeons were in second place and they were highest among the 6 groups of practitioners. The previous holder of the top rank among the Case Western Reserve graduates was the group of full-time medical school faculty; currently, full-time faculty members have dropped 3 ranks and are in fourth place. Climbing to the number 2 rank in the present study are the full-time hospital based personnel; this is a big jump upward for, in the previous study, they ranked in sixth place. There were small shifts upward for the internists (from fourth to third rank), for the general practitioners (from seventh to sixth rank), and for the pediatricians (from eighth to seventh rank). For the pediatricians, it was a major step out of last place. Obstetrician-gynecologists remained the same, in fifth rank, in both studies. Most surprising of all in this study was the psychiatrists' drop of 5 places from ranking third (second only to the general surgeons among the practitioners) to last or eighth place. Unfortunately, there were no job satisfaction data from the previous study with which to compare rankings for the women physicians. In the present study (1956-65 graduates) the women physicians' mean score of 77.06 would cause them to rank high--between the full-time hospital based personnel in second place and the internists in third.

DISCUSSION

What can be some of the factors behind these shifts in rank on job satisfaction? It is too soon to have in hand all of the data from the analyses of the many questions in the interview, so an attempt will be made to enumerate a few of the items that have been named as being at least partially responsible. Clearly, for graduates in a period of time spanning some thirty years, general surgery is a very satisfying specialty. In surgery there is general recognition and status accorded by the public, the satisfaction of being able to see relatively quick results and avoid long-term care of the chronically ill, and incomes of surgeons are the highest among all physicians. General surgeons emerged at the top in our present study despite the fact, as one said, "....there was nothing in the revised program for surgeons."

It was hypothesized for this present career study that the pediatricians would not again be the lowest on job satisfaction due to one of the special features in the revised program in medical education, namely, the Family Study Program. This was the feature where the medical student, upon entry, was introduced to clinical science by being assigned as student-physician to follow a pregnant woman. The student observed the patient through the remainder of her pregnancy and was present at delivery. Later, the student followed the child and other members of the family in a continuing way for about two years. It was felt that this experience would enable the student to have a better understanding of what continuing well-baby-care was like and that those who still elected to go into the field would know more about what they were choosing. In the previous study, a major complaint by pediatricians was that their preparation and training were so unlike what they had to do in practice. The mean job satisfaction score of pediatricians did rise several points and their rank rose from eighth to seventh place. However, the one extremely low score on the job satisfaction test in the present study was obtained by a pediatrician. This score was some thirty points below the lowest score of any other physician in any other group. His score, rightfully, was included in the pediatrician mean score. But, if one arbitrarily removes this single score which so skews the statistics, the mean score for the pediatricians is 75.53 and their ranking jumps all the way to third place! Clearly, most of the pediatricians in the present study are more satisfied than previously. It would be fascinating to know if this is a national phenomenon for, in the past, many general sources have named pediatricians lowest in job satisfaction, as was the case on our own previous study.

The shift in rankings for the full-time medical school faculty is not so surprising nor so difficult to understand. At the time of the earlier study, it was suggested that full-time medical school faculty had three areas of potential satisfaction / dissatisfaction--their patients, their students, and their research. A physician could hope that, if things were not going well in one area, they might be proceeding well in the others. At the time of the present study, there is or has been perceived difficulty in all three. Obtaining funds for research has become more difficult and considerable time must be spent, often fruitlessly, in writing proposals. Since the student rebellions of the early 1970's, frequently there has been less satisfaction in teaching. Time devoted to preparation for classroom teaching has brought negative feedback many times or the even more discouraging non-attendance at lectures. Instead of freedom to practice as much or little as one wished, as was formerly the usual situation, there is now pressure to practice more. Proceeds are now often pooled within a departmental group practice, and from these funds, the full-time medical school faculty member is now paid a pre-determined proportion for his practice activities. The strong financial squeeze at practically every medical school has plagued the life of full-time faculty members and created new dissatisfactions.

The most surprising result was the psychiatrists' drop in rank from third to eighth place, the least satisfied of all groups today. This had not been predicted at the time the present study was planned, but psychiatrists themselves have offered some partial explanations. During the last ten or fifteen years the public image of the psychiatrist has fallen more than that of other physicians; psychiatrists' incomes have not kept up with inflation to the same extent as that of other specialists; there is new competition for psychotherapy patients from psychologists, social workers, and others; there exists some in-fighting among psychiatrists themselves as they compete for the smaller pool of individual psychotherapy patients; the trend is away from the individual psychotherapy or psychoanalysis which many psychiatrists were prepared to do and toward more involvement in community mental health projects, drug and alcoholism centers, suicide prevention boards. Finally, psychiatrists appear to be the most susceptible of all groups to one of the newer stresses physicians have reported,(5) that of the threat of physical harm from disgruntled patients or their families. During this 1956-65 period, psychiatry was the second most popular chosen specialty among our graduates. The program and environment placed heavy emphasis on psychiatry and it is probable that many factors were conducive to the choice of psychiatry as a specialty. It is possible that some people selected psychiatry for whom it was not the most auspicious choice. The one psychiatrist in the study who had left that field for an entirely different specialty admitted he could not tolerate the isolation of the solo practice of psychiatry. It is interesting to note that the scores of psychiatrists who are in the categories of full-time medical school faculty and full-time hospital-based personnel are higher than for the group of private practice psychiatrists. Psychiatry was a popular choice of many women physicians and the women psychiatrists obtained a job satisfaction mean score of 79.00 compared to the 77.06 mean for all women. The women physicians, in general, seem to have relatively high job satisfaction in spite of many of the problems and subtle or not-so-subtle discriminations they report they have encountered.(6)

CONCLUSIONS

These are some of the shifts in job satisfaction for physicians from two career studies in two decades. There are more data and other issues to be discussed which space does not permit here. We have had the opportunity to learn what are some of the recent changes in job satisfaction for different physican groups and have just begun to investigate further sources for greater or lesser satisfactions.

REFERENCES

1. Mawardi, Betty Hosmer, "The Case Western Reserve Revised Medical Education Program: A Long-Term Evaluation By Graduates." Presented at the meetings of the American Educational Research Association in Boston, Massachusetts, April 11, 1980.

2. Mawardi, Betty Hosmer, Physicians and Their Careers. Ann Arbor, Michigan: University Microfilms International, 1979.

3. Glick, Ruth, "A Study of the Career Patterns of Women Physicians Who Received the M.D. Degree from Western Reserve University School of Medicine Between 1932 and 1958." In Physicians and Their Careers. (Appendix E).

4. A. H. Brayfield and H. F. Rothe, "An Index of Job Satisfaction," Journal of Applied Psychology, Vol. 35, 1951, pp. 307, 311.

5. Mawardi, Betty Hosmer, "Satisfactions, Dissatisfactions, and Causes of Stress in Medical Practice," Journal of the American Medical Association, Vol. 241, No. 14, April 16, 1979, pp. 1483-1486.

6. Mawardi, Betty Hosmer, "Styles of Practice of Female Physicians," Proceedings of Sixteenth Annual Conference of Research in Medical Education, Washington, D.C.: Association of American Medical Colleges, 1977.

The author wishes to express her appreciation for discussions with Russell Weisman, Jr., M.D., and L. Douglas Lenkoski, M.D. She is also grateful for financial support from The Commonwealth Fund, The Cleveland Foundation, and the David L. and Harriet T. Simon Philanthropic Fund.

TABLE 1

JOB SATISFACTORY SCORE MEANS*

Eight Groups of C.W.R.U. Graduates in Both Career Studies

Physician Group	1956-65 Study Means**	1935-45 Study Means***
Practitioners:		
General Practitioners	74.00	72.94
Internists	75.31	75.12
General Surgeons	79.08	79.36
Psychiatrists	73.07	76.89
Pediatricians	73.13	69.85
Obstetrician-Gynecologists	74.29	74.26
Sub-total	74.69	74.82
Non-Practitioners:		
Full-Time Medical School Faculty	74.33	80.18
Full-Time Hospital-Based Physicians	77.86	73.87
Sub-total	76.03	76.54
Total	75.03	75.12
Physicians in 1956-65 Study Only:		
Women Physicians	77.06	

Higher score = greater satisfaction

TABLE 2

JOB SATISFACTION RANKS*

Eight Groups of C.W.R.U. Graduates in Both Career Studies

Physician Group	1956-65 Study Rank**	1935-45 Study Rank***	Change in Rank
Practitioners:			
General Practitioners	6	7	+1
Internists	3	4	+1
General Surgeons	1	2	+1
Psychiatrists	8	3	-5
Pediatricians	7	8	+1
Obstetrician-Gynecologists	5	5	0
Non-Practitioners:			
Full-Time Medical School Faculty	4	1	-3
Full-Time Hospital-Based Physicians	2	6	+4
Physicians in 1956-65 Study Only:			
Women Physicians	2.5 (between full-time hospital in second and internists in third rank)		

*Measured by modified Brayfield Job Satisfaction Blank.
**Based on 142 books returned to date.
***1935-45 means and rankings are reprinted from Mawardi, Physicians and Their Careers, p. 103.

PRECIS

TEACHING CLINICAL SKILLS

MODERATOR: Edward Coppola, M.D.
Michigan State University, College of Medicine

INFORMATION MAPPING IN INTRODUCTION TO CLINICAL MEDICINE

This study investigated the influence of a highly structured (information mapping) format for presenting print information on the general screening physical examination to second-year medical students in an introduction to clinical medicine course. Results indicated significant attitudinal differences but no difference in performance on the final examination between students using an information mapped syllabus and those who received a conventional, prose formatted syllabus.

PERCEPTIONS OF STUDENT-PATIENT RELATIONS

Input from 49 patients about the interpersonal skills of second year medical students shows general satisfaction and that patients could be a source of feedback in the training of students. Areas patients identified as important were caring, listening, conscientiousness, decisiveness and gentleness on the part of the students.

A COMPARISON OF STRUCTURED AND SELF-DIRECTED APPROACHES TO TEACHING INTERVIEWING AND INTERPERSONAL SKILLS TO PEDIATRIC RESIDENTS

The paper summarizes a study which was conducted in 1979 comparing the effectiveness of the two most often used methods of instruction, structured or didactic and self-directed in teaching interviewing and interpersonal skills to first year pediatric residents at Children's Hospital National Medical Center in Washington, D.C.

TEACHING MEDICAL INTERVIEWING SKILLS: A COMPARISON OF MEDICAL AND NON-MEDICAL TUTORS

In a prospective, controlled study of a medical interviewing course, medical and non-medical (psychologist and social worker) tutors are compared for differences in educational process and outcome variables. Non-medical teachers were found to be at least as effective as medical teachers, particularly for facilitating interpersonal skills.

INFORMATION MAPPING IN INTRODUCTION TO CLINICAL MEDICINE

Emil Petrusa, Ph.D. (The University of Texas Medical Branch at Galveston), Paula K. Horvatich, Ph.D. (The University of Michigan) and James C. Guckian, M.D. (The University of Texas Medical Branch at Galveston)

INTRODUCTION

The general screening physical examination emphasizing normal physical manifestations is taught in Phase I of the Introduction to Clinical Medicine (ICM) course at the University of Texas Medical Branch at Galveston. Over the years ICM has received highly favorable ratings and has maintained its popularity with students. Only one feature of the course has received poor ratings; students have consistently indicated dissatisfaction with the course syllabus. The ICM syllabus is a compendium of handout materials (lecture outlines, descriptive narrations, diagrams, etc.) provided by individual lecturers for the eight examination segments which they present (screening physical exam in general, head and neck, eyes, chest/lung/back/breast, heart, abdomen and inguinal areas, pelvic, and neurological examinations). The actual materials for each topic vary considerably in terms of style and format. Students recommended that a more comprehensive and standardized format be developed for the syllabus.

Acting on the students' recommendation, the ICM Course Committee decided to develop a syllabus written in a systematic, organized, and appropriately sequenced manner based on principles believed to enhance learning from prose material. Information mapping is a unique, comprehensive design and presentation technology for developing print materials that has been used extensively for preparing technical communications in business and industry (Horn, 1975). It was derived from a synthesis of research findings on types of learning, instructional strategies, information processing, and graphic design Thiagarajan, 1977). Information mapping utilizes a system of principles for categorizing, writing, interrelating, sequencing, and presenting graphically information for learning and reference (Horn, 1974).

The conventional paragraph of prose is replaced by a series of carefully defined, functionally labeled blocks of information. These information blocks, are then compiled to form information map of various types. Each type of information map (procedure, process, concept, etc.) looks different on the printed page because it is designed to serve a particular purpose. In conventional learning materials, paragraphs look the same but may present learners with entirely different learning tasks. Moreover, because an information map contains functional labels, a learner can scan a page rapidly and locate readily a specific piece of information. In this manner the information mapping format facilitates both initial learning and review.

Information mapping's uniform set of principles and rules makes the task of producing instructional materials from the content submitted by subject

---

Send requests for reprints to Emil Petrusa, Ph.D., 409 Clinical Science Building, The University of Texas Medical Branch, Galveston, TX 77550.

72

matter experts, faster and easier. The modular nature of the information mapping format means that changes, rewrites, and updates can be accomplished without total reconstruction of the final product.

Because information mapping is especially suited for documenting procedures, it seemed an appropriate strategy to use for the presentation of the general screening physical examination. Each of the eight chapters in the conventional syllabus were given to an instructional designer who rewrote them in the information mapping format. The information mapped syllabus was derived from the conventional syllabus so that the actual content of the two was identical. Figure 1 provides an example of the syllabus formats for comparison purposes.

The attractiveness of information mapping for enhancing the organization, sequencing and readability of a syllabus, and its suitability for the presentation of the relatively technical procedures of the screening physical examination stimulated the preliminary investigation described in this paper. The purpose of the study was to compare the effect of an information mapped syllabus on student learning with that of a conventional syllabus.

METHOD

Subjects - The second year medical class (N=202) at the University of Texas Medical Branch at Galveston was divided into two groups by the Student Personnel Office. From an alphabetical list of the students' names, every other student was assigned to the same group. Although assignment to groups was not random, lack of significant differences on other potentially influential variables (age, sex, and average grades for freshman year in medical school) indicated the procedure approximated random assignment. One group of students was given the information mapped syllabus (MAP) while the other group received the conventional prose formatted syllabus (CONV).

Procedure - On the first day of class all students were oriented to the organization and grading procedures of the ICM course. Students were given no special instructions or orientation regarding the specific syllabus they received. They were told only that the two versions were being used for comparative purposes and that the content of each syllabus was the same. Students were asked not to share the syllabus they received with someone from the other group. All students proceeded through the remainder of the ICM course as usual.

On the last day of the ICM course all students completed the final examination. Examination questions were based on the material in the syllabus and were selected from an item bank developed for ICM over a period of several years. Test item security had been maintained from year to year. The final examination consisted of 109 objective type items (45 multiple choice, 30 true/false, and 25 matching). Scores were the percent of items answered correctly.

Students also provided feedback about the ICM course and syllabus. With the exception of identifying to which section (MAP or CONV) each student was assigned, two evaluation forms were completed anonymously. For the ICM Course Evaluation students rated the overall effectiveness of the course and various other aspects of the course on a four-point scale of Poor, Fair, Good, and Excellent. This form had been used for several years to evaluate ICM. The response rate for the course evaluation from the MAP group was 80 percent and 76 percent for the CONV group. A new form was developed to assess specific aspects of the syllabi. For the ICM Syllabus Evaluation, students rated the syllabus they received on (1) ease of reading, (2) organization, (3) sequence of information, (4) ease of locating information for review, (5) ease of studying, and (6) interesting information using a five-point scale of Poor, Fair, Good, Very Good, and Excellent. Students also estimated the average time spent reading the syllabus per week: 30 minutes or less, 45, 60, 75, 90, 105, 135 and 150 minutes or more. Response rates for this evaluation form were 74 percent for MAP, and 69 percent for CONV.

Responses to items on the ICM course and syllabus evaluations were analyzed by chi square tests. The number of students who rated an item in the upper two categories were compared with those rating the item in any of the lower categories. Analysis of the estimated weekly time spent reading the text also was analyzed by chi square with the separation occurring at 60 minutes per week. Differences between the groups on the final exam was assessed with a *t* test.

RESULTS

No significant difference was found between the MAP and CONV groups in performance on the final examination ($t=0.931$, $df=200$). The distribution of the MAP group had a range of 69 to 98, a mean of 87.67 and a standard deviation of 6.00. The CONV group was very similar with a range of 72 to 97, a mean of 88.39 and a standard deviation of 4.58. Kudar-Richardson 20, a statistical estimation of test reliability was .79 for the MAP group and .68 for the CONV group.

The ICM course received favorable evaluations from both groups of students. Ratings on the two questions from the ICM Course Evaluation form that are particularly relevant to the present study are included as the first two items in Table I. The ratings given by the MAP group were not significantly different from the ratings given by the CONV group.

Regarding the ICM Syllabus Evaluation, the information mapped syllabus received significantly higher ratings on all dimensions than did the conventional syllabus (see Table 1). Although the content of the syllabi was identical, students in the MAP group rated their syllabus significantly more "interesting" than the CONV students rated their syllabus. There was a significant difference in the proportion of students from each group who spent more than 60 minutes per week reading the syllabus ($x^2=6.59$, $df=1$, $p<.02$). Thirty-eight percent of those reading the information mapped syllabus estimated that they spent more than one hour per week studying from it, while only 25 percent of the students in the conventional group studied from their syllabus more than an hour per week.

DISCUSSION

The information mapping format did not produce a significant difference in student learning, at least as assessed by the final examination. Students in both the MAP and CONV groups performed well on the examination; 48 students in each group correctly answered 90 percent or more of the test items. Although the examination was fairly reliable, many items had a low discrimination index indicating that students who scored low and those who scored high both tended to answer those test items correctly. Another explanation for the lack of a significant difference is that the course syllabus was not the only means for learning about the general screening physical examination. Students were provided with the opportunity to attend lectures and view videocassettes on the general screening physical examination. There were eight lectures and eight videocassette programs that corresponded to each of the chapters in the syllabus. These learning experiences realistically could have reinforced what students had read thus making any differential formatting effect between the two syllabi negligible.

Students who used the information mapped syllabus rated the syllabus significantly more favorably on several dimensions than students who used the conventional syllabus. In reference to the ICM Course Evaluation, the two groups did not differ in their ratings of the overall effectiveness of the course or of the syllabus in general. Therefore, the differences obtained on syllabus ratings probably were not the result of a Hawthorne effect. The attitudinal differences seem to be related to qualities of the information mapping format.

Although students in the MAP group rated the syllabus high on readability, they report spending more time reading from it; significantly more so than the CONV students spent. This result is an apparent contradiction as information mapping is supposed to facilitate initial learning and review. The evaluation of reading time did not distinguish between the time spent in initial study and time devoted to review. Perhaps because the information mapping format was easier to review, MAP students spent more time doing so. Whatever the reason, this finding is somewhat disturbing for the strongest feature of information mapping is the possibility it holds for increasing efficiency in the learning situation.

CONCLUSION AND IMPLICATIONS

The study reported investigated the influence of a highly structured (information mapping) format for presenting printed information on the general screening physical examination to second-year medical students. Results indicated significant attitudinal differences but no difference in performance on the final examination between students using an information mapped syllabus and those who received a conventional, prose formatted syllabus. The lack of difference on final examination performance was attributed to low discrimination indices of test items, and confounding effects of lectures and videotape programs. It also may be that format of print materials alone is not a powerful enough variable to effect final examination performance, especially with medical students who are a highly motivated and proficient group of learners.

The significantly favorable student attitudes obtained support the reported advantages of an information mapping format for enhancing the organization, sequencing, readability, and review of print materials. Because medical content is complex, comprehensive, and constantly increasing in volume, it seems to be an ideal recipient of the benefits that a presentation technology like information mapping can offer. The potential of information mapping for the presentation and learning of medical content warrants further investigation.

REFERENCES

Horn, R. Information mapping: New tool to overcome the paper mountain. Educational Technology, 1974, 14, 5-8.

Horn, R. More about information mapping. Training, 1975, 12, 36-38.

Thiagarajan, S. Information mapping: A breakthrough in print format. Audiovisual Instruction, 1977, 22, 16-17.

Table I: Students' Ratings of Course and Syllabus

Variable	Number of Students Rating: Excellent and Very Good	Number of Students Rating: Good, Fair and Poor	Chi Square
Overall Course Effectiveness			
MAP	69	11	.12
CONV	66		
Syllabus (in general)			
MAP	75	4	1.20
CONV	63	8	
Ease of Reading			
MAP	47	27	10.67**
CONV	24	45	
Organization			
MAP	50	24	14.09**
CONV	24	45	
Sequence of Information			
MAP	46	28	12.19**
CONV	21	46	
Ease of Locating Information			
MAP	37	36	12.95**
CONV	14	55	
Ease of Studying			
MAP	42	30	11.16**
CONV	20	49	
Interesting Information			
MAP	46	28	3.60*
CONV	31	38	

Note: Chi square analyses based on df=1 with correction for continuity

* $p<.05$, one tailed ** $p<.001$ one tailed

Figure 1

Conventional	Information Mapped

AUSCULTATION

Following inspection, the examiner should perform auscultation of the abdomen. Palpation and percussion usually follow inspection in the examiniation of the other body systems. This change in the order of examination is suggested because the auscultatory findings may be altered by any manipulation of the abdominal wall. Consequently percussion and palpation, which may increase or decrease peristaltic sounds, are deferred until auscultation has been completed.

The diaphragm of the stethoscope should be placed lightly against the abdominal wall in order to avoid artifacts resulting from friction and compression of vessels. First, one should listen to the sounds produced by intestinal peristalsis. In the normal abdomen, bowel sounds are always present. These gurgling noises are normally highly variable in their frequency, intensity, location, and pitch. This wide variability is a result of the normal physiological activity of the bowel. The sounds at a particular site are dependent upon bowel contents and bowel motility which are ever changing. These sounds are difficult to describe and are best appreciated from the experience of listening to the abdomen of many normal individuals. Also, this exercise permits understanding of the normal wide variability of bowel sounds, which may be interpreted as evidence of disease by the inexperienced examiner. The recording of "lightly hypoactive" bowel sounds by the student usually represents an attempt to give significance to a normal finding.

Two abnormalities of the bowel sounds are significant: The virtual absence of peristaltic "rush." The absence of any sound heard after several minutes of continuous auscultation ordinarily represent the immobile bowel of peritonitis or paralytic ileus, (i.e., hypokalemia-induced surgical manipulation with reflex ileus). In contrast, when loud, gurgling borborygmi appear, it may be assumed that intestinal motility is increased or that intestinal content are being squeezed through a stenotic area. Simple hyperperistalsis in the non-obstructed bowel has many reasons (emotional tension, diarrhea, blood in the G.I. tract). It is the specific character of the peristaltic sound in stenosis or obstruction that differentiates them from those coming from a stimulated, but non-obstructed bowel. Increased sounds with a characteristic loud rushing, high-pitched tinkling quality often occur in mechanical intestinal obstruction. Most characteristic of the borborygmi heard in bowel is their accentuation during the waves of paroxysm and pain. These are caused by distention of the bowel and increased peristaltic activity proximal to the site of the obstruction.

AUSCULTATION - BOWEL SOUNDS

Change in Order of Exam
Auscultation follows inspection in the abdominal exam, while palpation and percussion usually follow inspection of other body systems. The order of the exam is changed because ausculatory findings may be altered by any manipulation of the abdominal wall. Thus, palpation and percussion, which may increase or decrease peristaltic sounds, are performed after auscultation is completed.

Normal Bowel Sounds
In the normal abdomen, bowel sounds produced by intestinal peristalsis are always present. Auscultation of the abdomen is employed to determine the differences between normal and abnormal bowel sounds.

Differences in Bowel Sounds
Intestinal peristalsis produces gurgling noises which are highly variable in:
- frequency
- intensity
- location
- pitch.

The wide variability of bowel sounds is the result of normal physiological activity of the bowel. Bowel sounds in a particular site are dependent upon ever changing bowel contents and motility.

Note to Beginners
Normal bowel sounds are difficult to describe. The experience of listening to many normal abdomens is the best exercise to understand the wide variability of normal bowel sounds, which may be interpreted as evidence of disease by the inexperienced examiner.

EXAMPLE: The recording of "slightly hypoactive" bowel by the student usually represents an attempt to give significance to a normal finding.

Procedure

STEP	ACTION	COMMENT
1	place diaphragm of stethoscope lightly against abdominal wall	light pressure is necessary in order to avoid artifacts resulting from friction and compression of vessels
2	listen for normal intestinal peristalsis	wide variability of sounds possible
3	listen for 2 significant abnormalities	peristaltic "silence" and "rush" (see below)

Indications

ABNORMALITIES	DESCRIPTION	PROBABLE CAUSE
"silence"	virtual absence of peristaltic sounds for few minutes	immobile bowel of peritonitis OR paralytic ileus (i.e. hypokalemia-induced, surgical manipulation with reflex ileus)
"rush", i.e. loud, gurgling borborygmi	absence of specific traits noted below and increased frequency of peristaltic sounds	increased intestinal motility (i.e. simple hyperperistalsis in non-constricted bowel) possibly due to: • emotion tension • diarrhea • blood in G.I. tract
	increase of loud-rushing, high-pitched tinkling accentuated during waves of paroxysm & cramping pain	intestinal contents' being squeezed through a stenotic (constricted) area, i.e., intestinal obstruction

PERCEPTIONS OF STUDENT-PATIENT RELATIONS

Margaret E. Uguroglu, M.Ed., Center for Educational Development
University of Illinois at the Medical Center

Rodney Nelson, M.D. and Charles Kanakis, M.D., Abraham Lincoln School of Medicine
University of Illinois at the Medical Center

Review of Literature

"Creation of evaluation instruments demands broad involvement of all those affected by the use of such tools" (Thompson, 1969, p. 561), yet input from the patient seldom is considered in the construction of the tool or in the assessment of student performance. Attempts have been made to study patient satisfaction or compliance (Wooley, 1978; Hayes, 1978; Dimatteo, 1979) and practicing physician-patient relationships (Stillman, 1978; Dimatteo, 1979); however, the use of patients in the diagnostic assessment of student performance has generally only been with trained patients who also help in the teaching of interviewing or physical examination skills. (Scott, 1975; Zakus, 1976, Stillman, 1977).

The need for continued input from the patient in defining the constructs upon which students might be evaluated in student-patient relationships arises, in part, from the lack of emphasis in this area from admission requirements, current curriculums and the teaching methods of faculty (Zakus, 1976; Daggett, 1979). Ward and Stein (1975) in their study to reduce emotional distance take the position that too much teaching emphasis has been given to the content of the interview and not enough to the process. When observing interviewing skills, the content is referred to as the information obtained from the patient, while the process includes the interviewer's attitudes and interpersonal skills (Stillman and Sabers, 1978). Generally, content refers to what is done and process refers to how it is done.

Dimmateo's (1979) review of the literature concludes that rapport between the physician and the patient is medically important yet Friedman (1979) charges that interpersonal relations in health care as a field of inquiry is still in its infancy. The reasons are many. For example, in studying student-patient relations, toward the end of the sophomore year, students usually complete an examination which contains sections on history taking, physical examination skills, record keeping and student-patient relations to diagnostically assess their skills. Inherent theoretical and practical problems exist in such an evaluation. Some include variations in the personalities of patients, observers and students; complexity of patient history; different hospital settings; and observers rating different numbers of students (Stillman and Ruggill, 1978; Smith, 1979). Since the most knowledgeable judge, at least of the student-patient relationship, would appear to be the patient; an interview to elicit perceptions by the patient may prove helpful. The patient would be able to point out specifically the strengths or weaknesses he/she felt during the encounter with the student. After the proper amount of development of a system of feedback, regular patients could become an essential element in the feedback to students about their interpersonal relations.

Reprint requests should be sent to Margaret Uguroglu, CED, 808 South Wood, Chicago, Illinois 60613

Purpose

The major purpose of this pilot study is to see whether there are any behaviors that patients feel are important which have not been included in the instruction and evaluation after limited clinical experience. A second purpose is to determine if there is a significant difference between the evaluations of students by the observers and the evaluations by the patients involved in the examination in regard to student-patient relations. Finally, this exploratory survey seeks information from the process of interviewing patients about students that might help in the development of an input system from patients.

Methodology

Procedures/Sample

After four months of participation in the Phase I curriculum, students are required to pass a clinical skills examination. It takes place in a one-student, one-patient and one-observer hospital setting. The students have two hours to complete an interview and physical examination. Between one and twenty-six hours later, the same interviewer used a questionnaire to discuss with patients their perceptions about the examination. All patients, while generally randomly selected, were to be physically and mentally able to participate in the exam, able to give a history and not to have had unduly long past history of illness or hospitalization.

Group I included a sample of 34 patients from a university or veterans administration hospital in the inner city. Thirty were male and 4 female with a split of 16 whites and 18 blacks. Only a visual estimate of age was taken and 90% were in the 35 to 70 age range with more probably concentrated around 45 to 65 years. Group II included 15 patients from a community general hospital on the edge of the city. Anonymity was maintained in this second group, where patients after reading a letter explaining the purpose and details of the survey, gave written answers on the questionnaire immediately after their examination.

Similar to the unidentified group, patients personally interviewed were also told that the purpose of the questions was to improve the examination and to help students in the future. The interviewer stressed that the answers were confidential and would not affect the student's grade or their own medical treatment. Patients were given the option to withdraw from the study, but none chose to do so.

Instruments:

After an informal discussion of the type of information appropriate to the study, the authors independently formulated questions for the interview based upon past literature and their own experience. The questions were then consolidated and further revised by other educational consultants. A final 22 item questionnaire included three parts in the following order: a) open-ended questions, b) closed questions and c) closed questions correlating to the items on the observation form used by the attending physicians who were the observers. While the closed questions could be answered "yes" or "no," the interviewer had a five point rating scale and recorded the patient's responses as follows: 5 points = definite yes, 4 points = yes with hesitation, 3 points = not sure, 2 points = no with hesitation, and 1 point = definite no. A preliminary field test of the patient questionnaire indicated no major problem.

The attending physicians used an observation form which included seven questions about student-patient relations on an outstanding, satisfactory or un-satisfactory rating scale.

Frequency counts, correlation coefficients, and tests of significance were completed on the closed questions as well as a content analysis of the open ended items.

Results/Discussion

With input from a total sample of 49 student-patient interviews, the results of this study show that these patients were satisfied with second year medical students' initial physical examination and interviewing attempts. For those areas included in the questionnaire with a closed question, Table I shows the percent of patients generally rating students very positively except for nervousness and the use of the patient's name.

Table 1

PATIENT RATINGS OF STUDENTS

Descriptions	Percent Rating 5	4	3	2	1	Mean Points
Did you have enough chance to explain your problem?	85	11	2	0	2	4.7
Were you asked about any allergies?	92	4	2	0	2	4.8
Were you asked whether you smoke?	86	2	2	0	8	4.6
Were you asked whether you drink alcohol?	88	0	2	0	4	4.7
Were you asked how much alcohol you drink?	79	0	7	2	12	4.3
Student-doctor did not seem nervous.*	50	21	8	13	9	3.9
Did you understand what the student-doctor was saying to you?	85	8	2	2	2	4.7
Did you have a chance to ask the questions you wanted?	93	5	3	0	0	4.8
Did the student-doctor tell you his name?	91	2	7	0	0	4.8
Did you feel at ease during the exam?	79	17	4	0	0	4.8
Did you feel your privacy was respected?	96	4	0	0	0	5.0
Questions did not embarass.*	96	2	0	0	0	5.0
Did the student-doctor call you by name?	64	6	11	2	17	3.9

	Percent Yes	No
Would you ask for this student-doctor again?	87	13
Did you notice what the student-doctor wore?	78	19

*Question reversed when asked. Weights have been reversed.

Gutek (1978) warns, however, that one reason for distrusting measures of satisfaction is simply that people seem to be satisfied with everything that social scientists ask them about. He cites Taylor as saying that a satisfied person can be one who cannot think of anything at the moment that he/she would like to change. This is consistent with the data from the open ended question that asked patients what they would want to change in another examination, and the majority replied, "nothing." However, when asked to describe what the student's weaknesses were, about one fourth of the patients had definite ideas. Some said the students were too slow, repeated themselves, "went by the book" which meant they kept looking at their notes, gave an uncomfortable examination of the ear, and were not personal enough. An area which one might expect which was specifically described was inexperience. The patients sensed an indecisiveness or hesitancy on the part of the students.

80

Positive attitudes and actions of the students which patients listed outnumbered the negative ones four to one. These included caring, listening, easy going attitude, friendliness, politeness, thoroughness, conscientiousness, and competence. Patients on a whole were not embarrassed by questions and felt their privacy was respected. Also, an interesting quality which this sample mentioned repeatedly was gentleness. In combining the responses from the two open-ended questions, it appears that these patients were looking for competence, caring, decisiveness yet gentleness in their contact with medical personnel.

For the closed questions, patients did vary on their response to whether or not the student was nervous, and this indicates an area to explore in future studies. For example, follow up questions such as, "What did the student do that you thought he/she was nervous?" "Why do you think the student was nervous?" Several did say that having the attending physician there observing would make any student uncomfortable. A final point from Gutek (1978) is that people respond to questions about satisfaction in light of what they have already experienced. Since the students seemed to exhibit those skills patients have observed with their regular physicians, patients generally rated them favorably.

Because of the limited variance in both the patients' ratings of the students and the attendings' ratings of the students, many strong conclusions about the correlations between the two ratings cannot be made. The observation instrument used by the attendings included only a three-point scale and most students were given a middle rating. The majority of correlations for the 34 patients in the identified Group I, therefore, were not significant. The best item which asked whether the patient felt at ease during the exam did correlate frequently. Even with little variance, answers to this question correlated positively ($p < .05$) with 11 of the total 31 ratings which the attendings gave the students. The range was +.32 to +.39. Four additional items which correlated with this item of ease at $p < .01$ were: a) elicits a complete past medical history (+.44); b) applies skillfully the fundamental techniques of examination (+.41); c) restrains from assigning unwarranted medical terms to patient descriptions in the written record (+.64); and d) does not omit major physical findings from the written record (+.59). From the 31 items which the attendings used, these four appear to be key points and patients' ease correlated most significantly with them. Interestingly, the last two correlations concerning the write-up had the highest relationship. Using a one-way causal interpretation this may suggest that students who were totally concerned with listening to their patients and therefore able to record their findings accurately, put patients at ease the most.

Group I, the personally interviewed sample from the inner-city hospitals generally did not differ from the unidentified Group II sample from the community general hospital. The differences that arose between groups are presented only for speculation, since they resulted from exploratory analysis rather than as a specific feature designed in the study. It is impossible to determine whether the differences which did result occurred because of the personal interview or a difference in population.

One area which was different on a chi square test of significance ($p < .05$) was that of appearance. Twenty-eight percent of Group I did not notice what the student doctor wore while one hundred percent of Group II did. This association may reflect the possibility that more of the patients at the inner-city hospitals were from a lower socio economic group and do not pay as much attention to the possibly more superficial aspects of medical care. From the 37 patients who did notice appearance, when asked to elaborate, only two had negative judgments that they thought a student could have been a little neater. The other area which was significantly different ($p < .01$) between groups involved the use of the patient's name. When those from Group I were asked whether the student-doctor called them by name, 84% replied "yes", while only 40%

in Group II responded positively. A possibility for some of the difference may be that the interviewer used a follow-up probe if respondents seemed puzzled by the question. Seventy percent of the patients stated that the student was either introduced to them or told them their names, but when aksed what it was, only one third were able to recall it or some name close to it. There was no significant correlation between students whose patients remembered their names and the ratings they received by attendings.

Since age was only an estimate and most of the patients grouped into one general range, no analysis seemed appropriate for this classification. Also, with only four identified females, sex could not be used as a variable. From the identified sample, however, the proportion of blacks and whites divided into approximately equal size, but analysis of the responses by race showed no differences.

Some of the limitations of this study appear in the previous discussion. Additionally, one cannot dismiss the halo effect that if a patient liked a student from the beginning, their ratings would all tend to be high. Secondly, the representativeness of the sample prevents generalizations to other populations. This exploratory study focused on the process involved in gathering data from patients about students. Suggestions in methodology for the future would include the following.

a) Two to four hours after the examination appears to be the ideal time for interviewing since immediately after, the patient begins to show signs of questioning overload. By the next day, some of their observations have faded.

b) De-emphasize the use of the word student in the interview, and possibly substitute "person," since patients may be more inclined to overlook negative behaviors in a student.

c) Patients were always told that the interview would only take five to ten minutes. This worked effectively and, in fact, the interview could probably be lengthened to 10 to 15 minutes if a new study warrants, since patients did not seem unduly inconvenienced and some were even surprised by the brevity of the current interview. A final open question of, "Is there anything else you would like to tell me about?" would encourage input from the patient which the interviewer may have missed.

d) No patient was awakened for the study. Although several had to be eliminated from the sample because they were asleep each time the interviewer attempted to see them, it was felt that they should not be disturbed. Also some may have reacted in their answers to the intrusion rather than to the examination by the student.

Conclusion

Only a few of the correlations between examiner ratings and patient ratings of the students were significant. The question which positively correlated most frequently with examiners' ratings was whether the student put the patient at ease. Patients discriminated between those students who were nervous and those who were not. Further exploration into these two areas is suggested and may yield more concrete suggestions on how students could improve upon their physical examination and interviewing skills. If patients are not inconvenienced, they are very cooperative in giving feedback about student performance. The areas patients felt were important were caring, listening, conscientiousness, decisiveness and gentleness. Generally, patients were satisfied with beginning medical students first physical examination and interviewing attempts.

REFERENCES

Daggett, C.J., Cassie, J.M. & Collins, G.F. Research on clinical teaching. Review of Educational Research, 1979, 49, 151-169.

DiMatteo, M.R. A social-psychological analysis of physician-patient rapport: toward a science of the art of medicine. Journal of Social Issues, 1979, 35, 12-33.

Friedman, H.S. & DiMatteo, M.R. Health care as an interpersonal process. Journal of Social Issues, 1979, 35, 1-11.

Grayson, M., Nugent, C., and Oken, S.L. A systematic and comprehensive approach to teaching and evaluating interpersonal skills. Journal of Medical Education, 1977, 52, 906-913.

Gutek, B.A. Strategies for studying client satisfaction. Journal of Social Issues, 1978, 34, 44-56.

Hayes, D.M. Perceptions of cancer care by professionals and nonprofessionals. Evaluation and the Health Professions, 1978, 1, 29-56.

Kennell, J.H., Tempio, C.R., & Wile, M.Z. Self-evaluation by first-year medical students in a clinical science programme. British Journal of Medical Education, 1973, 7, 230-236.

Scott, N.C., Donnelly, M.B., Hess, J.W. Changes in interviewing styles of medical students. Journal of Medical Education, 1975, 50, 1124-1126.

Smith, D.U. Estimating the reliability of clinical ratings: an alternative approach. Paper presented at the meeting of American Educational Research Association, San Francisco, April 1979.

Stillman, P.L., Sabers, D.L. & Redfield, D.L. Use of trained mothers to teach interviewing skills to first year medical students: a follow-up study. Pediatrics, 1977, 60, 165-169.

Stillman, P.L. & Sabers, D.L. Using a competency-based program to assess interviewing skills of pediatric house staff. Journal of Medical Education, 1978, 53, 493-496.

Stillman, P.L., Ruggill, J.S., & Sabers, D. The use of practical instructors to evaluate a complete physical examination. Evaluation and the Health Professions, 1978, 1, 49-54.

Stillman, P.L. & Ruggill, J.S. The use of patient instructors (PIs) to teach and evaluate medical interviewing. In Paula L. Stillman, Annell Gooden, Jose F. Laguna and Paul J. Rutala (Eds.) The non-physician in medical education. Arizona: University of Arizona, 1978.

Taylor, S.E. Hospital patient behavior: reactance, helplessness or control? Journal of Social Issues, 1979, 35, 156-184.

Thompson, A.G. The shadow and substance of evaluating students in medical technology. American Journal of Medical Technology, 1969, 35, 561-567.

Ward, N.G. & Stein, L. Reducing emotional distance: A new method to teach interviewing skills. Journal of Medical Education, 1975, 50, 605-614.

Woolley, F.R., Kane, R.L., Hughes C.C. & Wright, D.D. The effects of doctor-patient communication on satisfaction and outcome of care. Social Science & Medicine, 1978, 12, 123-128.

Zakus, G.E., Hutter, M.J., Dungy, C.I., Moore, V.M. & Ott, J.E. Teaching interviewing for pediatrics. Journal of Medical Education, 1976, 51, 325-331.

A Comparison of Structured and Self-Directed Approaches to Teaching Interviewing and Interpersonal Skills to Pediatric Residents

Leslie S. Jewett, Ed.D.*

INTRODUCTION

The historical development of higher educational standards in medical education and societal pressures in the United States have produced concern with the teaching and learning technology in medical education. There is now a growing interest in the acquisition of knowledge as well as the actual skills required in medical education. George Miller (1968) gave recognition to this issue, the technology of teaching, when he found that he knew very little about the learning process with his medical students and "that what I did to my students as often impeded their learning as facilitated it." Miller finds that a chief problem in medical education is the conformity and passivity of students. Furthermore, the educational model in medicine develops an authoritarian atmosphere in which there is little opportunity for students and residents to gain skills of independent judgment. This is not in keeping with the professional role of a physician whose importance in health care resides not in the accumulation of data but in problem-solving. Miller suggests the importance of having the students more actively participate in the educational process instead of being passive recipients.

Although there now appears to be a growing interest in and recognition of the need to improve the educational technology in medical education, few studies have been made. At the medical school level, Bazuin and Yonke (1978) found that faculty taught in the manner in which they had been taught without formal training in educational procedures; lecturing rather than problem-solving was the primary teaching approach. A few studies of the teaching-learning process at the residency level have been made with the focus on residents' learning in specialized areas of interviewing skills (Farsad, et al., 1978), psychiatry (Goin, 1976), and family practice (Johnson, 1977). However, although the need for improved medical education is recognized, the literature search revealed that no study has been made comparing the structured (passive) versus the self-directed (active) approaches to learning.

In response to this issue of the effect of active versus passive participation of students on their learning in the educational process, a study was made in which these two different instructional technologies were used in teaching history-taking and interpersonal skills to first year pediatric residents. The purpose of this report is to describe the impact of each of these instructional approaches in a Parent Interview-Instructional Program designed to improve the interviewing and interpersonal skills of first year pediatric residents.

METHOD

This study, conducted at Children's Hospital National Medical Center (CHNMC) in Washington, D.C., involved all of the 18 first year pediatric residents and a special pediatric resident. Fourteen of the 18 residents participated in each of the three interview-feedback sessions while four residents and the special pediatric resident took part in the first two sessions only.

The goal of the Parent Interview-Instructional Program is to improve the

*Office of Medical Education, Children's Hospital National Medical Center, 111 Michigan Avenue, N.W., Washington, D.C. 20010

quality of pediatric medical care by increasing the residents' competency in interviewing and interpersonal skills. This program began in 1975 at CHNMC and has continued to provide training each year in interviewing and interpersonal skills for the first year pediatric residents (PL-1s). The program has three major objectives: 1) to provide instruction in history-taking to the PL-1s to enable them to record the patient's medical history more accurately and efficiently; 2) to provide feedback to the residents regarding their interpersonal skills to enhance their effectiveness in parent/physician interactions; and 3) to evaluate the effectiveness of the simulated parent interview-instructional session's techniques of instruction (active and passive participation) as a means of improving history-taking and interpersonal skills.

Procedures

The Parent Interview Program consisted of an orientation and three one-hour instructional sessions. The first session was a brief orientation to the program and the remaining three sessions, scheduled individually with each participant, consisted of three one-hour parent interview-instructional sessions. Each session was divided into three parts: 1) History-taking (15 minutes): The resident, while being videotaped, was allotted 15 minutes in which to elicit a pediatric history from a simulated mother who presented one of a number of histories typical of inpatient or outpatient problems; 2) Instruction (40 minutes): a) Checklists (10 minutes): Immediately following the interview, the resident filled out a true-false history checklist and scored it with a template. This provided immediate feedback concerning what percentage of available information (relevant historical data known by the simulated mother) was elicited. During this time the simulated mother also completed a ckecklist which identified her feelings during the interview. b) videotape (30 minutes): The resident and simulated mother-instructor reviewed the videotape of the interview, pausing to discuss aspects of the interaction which elicited or failed to elicit important information and feelings. 3) Evaluation (5 minutes): The resident and simulated mother-instructor completed an evalution form, rating the importance and effectiveness of the interview session objectives in providing instruction in: 1) efficient history-taking and 2) effective interpersonal skills. They commented upon positive aspects of the interchange and how the session might be improved.

The components of these sessions were essentially the same. However, the style of instruction was the independent variable with the dependent variables being the resident's performance on the history-taking true-false checklist, the impact on the mother's feelings during the interview and the resident's style of interaction. In the first method of instruction, structured, the resident was passive in the learning process; the instruction being a didactic and telling process. In the second method of instruction, self-directed, the resident actively participated in the learning process in which the instruction was conducted through inquiry.

Program Design

This program was designed to evaluate the impact of the style of instruction used in the Parent Interview-Instructional sessions on the PL-1s' ability to take efficient and interpersonally effective pediatric histories. Residents were assigned randomly to two groups, and the Pretest-Posttest Design was used. Although there was not a control group in this study, the research results of past programs at CHNMC indicate that it was the instructional program itself which had an impact on the residents' learning of interviewing and interpersonal skills while time or the residency training program had no effect on the residents' skills in this area (Jewett, et al., 1979). Thus, it was concluded from these past studies and the small number of residents in the present study that it was preferable to have two experimental groups and no control group in the 1979 Parent Interview-Instructional Program.

The 1979 Parent Interview-Instructional Program provided three simulated parent interview and instructional sessions for the first year pediatric residents during early May, early June and September 1979. Data was collected immediately following each interview but prior to each instructional session. The residents were randomly assigned to the two groups and an instructor subgroup within each group.

The program was provided to subgroups of three PL-1s scheduled in random order to the simulated mother-instructors. The mothers were typical, middle class, white urban mothers in their thirties. Three different case histories were learned by each simulated mother and the order with which these cases were presented to each resident over the three interviews was randomly assigned. These cases have been found to be reliable in their level of difficulty (F=1.72).

Thus, the difference in performance between the two groups of residents at the time of the second interview session reflects the impact of the type of instruction received by the residents one month earlier. Differences in residence performance at the time of the third interview reflect the impact of the style of instruction four months after the first instructional session. The t-test was used for statistical analysis.

RESULTS

1. Efficiency in History-taking

The first objective of the Parent Interview-Instructional Program was to provide instruction in history-taking to PL-1s so that more efficient history-taking abilities would be attained. The criterion used for assessing performance in this category was the amount of present and past history elicited by each resident during the interview session. The effect of the style of instruction: structured (passive) or self-directed (active) on the amount of present and past history collected during the interviews, as measured by the performance of the residents on the 50-item history checklist, was evaluated.

It was found that residents in both instructional groups elicited a similar proportion of a comprehensive present and past history during their first simulated parent interview (Table I). No significant differences were found between the two groups who then received different styles of instruction. Therefore, it appears that Group I and Group II, into which the residents were randomly distributed, can be considered comparable and the effect of the two styles of instruction, active and passive, can be assessed.

It can be seen one month later (interview 2), however, (Table I) that the first instructional session, which provided training in interview techniques and interpersonal skills, affected the residents' efficiency in obtaining a more comprehensive history only when the instructional session was conducted in a structured, didactic style. Thus, residents in Group I, who passively participated in a structured learning session, obtained significantly ($p < .03$) more past history and significantly ($p < .02$) more of the total history at the time of the second interview than did residents in Group II who actively participated in a self-directed learning session.

TABLE I: Residents' Mean Percentage Scores on 50-item History Checklist

INTERVIEW:	PRESENT HISTORY 1	PRESENT HISTORY 2	+	PAST HISTORY 1	PAST HISTORY 2	=	TOTAL HISTORY 1	TOTAL HISTORY 2
GROUP I RESIDENTS: (N=10)	64%	76%		45%	60%*		54%	68%**
%Gain:		19%			33%			26%
GROUP II RESIDENTS: (N=9)	65%	68%		47%	47%*		56%	57%**
%Gain:		5%			0			2%

*$t=2.32$; $s_1=3.302$, $s_2=3.232$; $p < .03$ **$t=2.52$; $s_1=4.667$, $s_2=4.555$; $p < .02$

The majority of the residents in both Group I and Group II participated in a third interview-instructional session three months following their second 15 minute pediatric interview-instructional session in month two of the program. Three residents in Group I and one resident in Group II were unable to return for a third interview for reasons of leaving the program or inavailability.

It was found at the time of this third interview, that residents in Group I, who had participated in two structured learning sessions, still elicited more comprehensive past and present histories and thus a total history than did residents in Group II who had actively participated in two self-directed learning sessions. Group I residents became more efficient at gathering information at the time of the third interview, particularly in the area of past history where they were able to elicit 72 percent of the available relevant information and obtain 74 percent of the total history. Group II residents also demonstrated some increase in their efficiency to elicit past history information, obtaining 57 percent of the past history and overall eliciting 65 percent of the total history. However, although Group II residents made some gains in their efficiency by the time of the third interview, Group I residents still surpassed Group II residents in their efficiency in obtaining relevant historical information in a pediatric interview; this was particularly true in the category of past history where Group I residents elicited 72 percent as compared to 57 percent elicited by Group II residents during the third interview. This was not found to be significant.

In summary, it was found that residents who participated in passive, teacher structured instructional sessions demonstrated significantly ($p < .02$) higher gains in their history-taking skills than did residents who participated in active, self-directed instructional sessions.

2. Effectiveness of Interpersonal Skills

The second objective of this program was to provide instruction in interpersonal skills to enhance the residents' effectiveness in parent-physician interactions. These skills were evaluated by studying the amount of positive and negative feelings in the mother which were elicited by the residents during the videotaped interview. The impact of the style of instruction, structured (passive) or self-directed (active) on the positivity of the mothers' feelings, as measured by the trust and feelings checklist, was evaluated.

The results of the first resident-mother history-taking interview indicated that the residents in both groups elicited a similar level of trust and feelings related to interpersonal effectiveness during the first interview which was prior to any instruction being given to the residents. The residents in both groups elicited moderately positive feelings of trust, reassurance, interdependence in and of listening to the mothers.

It was found that one month following the initial interview-instructional session, that residents in both groups showed similar gains in their effectiveness in interpersonal skills, being able to elicit more positive feelings (ranging from 3% to 20% increase). Mothers also reported that residents elicited fewer of the extreme negative feelings and more of the highly positive interpersonal feelings at the time of this second interview. However, there was [illegible] significant difference found in the interpersonal skills of residents with mothers between Group I and Group II. Thus, the style of instruction does not appear to have a significant impact on the residents' learning of interpersonal skills.

In a follow-up session four months after the initial interview-instructional session and three months after the second similar session, it was found that residents who had participated in structured-passive learning sessions, showed [illegible] substantial improvement in their interpersonal skills with mothers than did residents who had participated in self-directed (active) learning sessions. [illegible] there was not a significant difference between the two groups of residents at the time of the third interview, which may be attributed to the smaller number participating, Group I residents made significant gains in their abilities to elicit more positive and fewer negative feelings in mothers during the third interview in almost every category of interpersonal skills (Table II). This was not true for Group II residents who showed mostly only slight gains with a significant gain in listening skills with the mothers. In conclusion, the results of this study indicate that residents participating in structured learning sessions made many more significant gains in their interpersonal skills by the end of the program than did residents who participated in self-directed learning sessions.

TABLE II: Mean Range of Affect Experienced by Mothers with Residents in Interview 3 and Percentage Change from Interview 1.

		DISTRUST+- TRUST++		ANXIETY+- REASSURANCE++		NOT LISTENED TO+- LISTENED TO++		DEPENDENCE+- INTERDEPEDENCE++	
	Range:	Low	High	Low	High	Low	High	Low	High
GROUP I(N=6):	Interview 3:	9.8	11.0	8.3	10.7	9.5	11.0	7.3	10.[illegible]
	%Change from Int.1:	69%*	5%	73%**	16%	[illegible]9%***	18%****	41%*+	30%*++
GROUP II(N=8)	Interview 3:	7.5	10.[illegible]	5.9	10.0	8.5	10.8	4.9	9.5
	%Change from Int.1:	16%	5%	23%	0	58%*+++	26%	29%	4%

+1-5 on the 11 point scale

++7-11 on the 11 point scale

*t=2.70; s_1=2.3167, s_3=2.0412; $p < .02$

**t=2.96; s_1=2.6394, s_3=3.5024; $p < .05$

***t=5.59; s_1=2.1366, s_3=2.0736; $p < .01$

****t=3.95; s_1=1.0327, s_3=0.0 ; $p < .02$

*+t=2.70; s_1=2.6580, s_3=3.8297; $p < .03$

*++t=[illegible]; s_1=1.7607, s_3=0.8165; $p < .01$

*+++t=2.80; s_1=1.0198, s_3=2.[illegible]039; [illegible].01

3. Program Evaluation

At the end of each session, residents completed a program evaluation form evaluating the "importance" and "effectiveness" of the session in providing instruction in efficient data collection and in interpersonal skills. Residents in both groups gave similar ratings to the program, rating the program to be between "important" and "extremely important" and the "effectiveness" of the program to be slightly higher than "effective."

DISCUSSION

The results of this study suggest two major findings:

First, of the two methods of instruction which were tested, it appears that teacher-structured passive learning is significantly more effective than self-directed student-active learning as a technique in teaching history-taking skills to first year pediatric residents. The residents participating in structured learning sessions demonstrate the largest gain in their history-taking effectiveness after the first instructional session but still show some gains after a second interview-instructional session when there is a three month interval between the second and third session. In comparison, residents having self-directed learning show small gains in their history-taking effectiveness after both the first and second instructional sessions.

Secondly, although it was found that there is not likely to be a significant difference in structured and self-directed instructional approaches to teaching interpersonal skills, residents, who participated in structured learning sessions, tended to make many significant improvements in their interpersonal skills with mothers by eliciting more positive and fewer negative feelings in the mothers. In contrast, residents who participated in self-directed learning sessions, made only slight gains in their interpersonal skills and very few significant improvements in their interpersonal skills with the mothers.

A possible reason for the teacher-structured passive approach being more successful in teaching interviewing and interpersonal skills is that this is the traditional way in which medical students and residents are taught. This study and another study indicates, however, that residents do not appear to prefer one style of teaching over another. Residents in both instructional groups in this study gave similar ratings to the importance and effectiveness of the program. In another program in which these residents also participated, little difference was found on The Learning Preference Inventory between resident preferences for teacher structured learning and student structured learning (Goldberg, Greenberg and Jewett, unpublished). Thus, although residents appear to learn interviewing and interpersonal skills more effectively with teacher-structured passive learning, there is evidence that they do not prefer this teaching style to the student-directed style.

Before further conclusions can be made about teaching methodologies, more research is needed. The limitations of this study with its small number of participants and the paucity of research in teaching methodologies, indicate further study is necessary to adequately evaluate the impact of these two instructional approaches in both teaching interviewing and interpersonal skills and in other areas of medical education.

REFERENCES

1. Bazuin, C.H. and Yonke, A.M. Improvement of Teaching Skills in Clinical Settings. J. Med. Ed., 52:377-382, 1978.
2. Farsad, P., et al. Teaching Pediatric Interviewing Skills to House Officers. Ped., 61:382-388, 1978.
3. Goin, M.K. Teaching Dynamic Psychotherapy by Observation. Am. J. Psychotherapy, 30 (1):112-120, 1976.
4. Goldberg, R., Greenberg, L. and Jewett, L. Teaching Residents How to Teach. Unpublished study.
5. Jewett, L., Richardson, G. and Richardson, H. History-Taking and Interpersonal Skills Training for Pediatric Residents. Clinical Proceedings, 35 (6): 346-352, 1979.
6. Johnson, A.H. Developing Behavioral Science for a Family Practice Residency. J. Fam. Pract., 4:319-324, 1977.
7. Miller, G. Interview. The New Physician, Sept.: 17-22, 1968.

TEACHING MEDICAL INTERVIEWING SKILLS: A COMPARISON OF MEDICAL AND NON-MEDICAL TUTORS

JULIAN BIRD, M.R.C.P., M.R.C. PSYCH., CHRISTOPHER D. LORISH, Ph.D.
STEVEN COHEN-COLE, M.D., C. KIRK AVENT, M.D.

Introduction

This study compares the effectiveness of non-medical professional staff (psychologists and social workers) with that of internists and psychiatrists in teaching medical interviewing skills to medical students.

There are few reports of non-medical staff being invited to teach medical interviewing and reports that are available do not make rigorous comparisons (Carroll, 1980). Specially trained mothers in a pediatric clinic can give very useful feedback to medical students (Stillman et. al., 1976, 1977) (Helfer et. al, 1978). Senior medical students as teachers compare favorably with faculty in the eyes of the client students but do not produce superior interview skills as judged by observers (Barnes et. al, 1978).

If this paucity of reports does reflect actual lack of involvement by non-medical staff, it may mean that the restricted bio-medical model of the doctor's interview role, namely data gathering and instruction giving, is still widespread. In this model doctors are the only eligible teachers, being the experts on what data to gather and on what decisions to make.

In the last few years evidence has accumulated that a doctor who uses emotional support skills and compliance fostering skills will achieve better health outcomes (Engel, 1977). Non-medical staff, for example psychologists and social workers, may well be more expert than doctors in these areas. Futhermore, such skills would seem closely linked to the teaching process. For these reasons it was hypothesized that psychologists and social workers may be as or more effective than doctors as teachers of the medical interview process, if adequate guidelines as to biomedical content are available.

Method

Personnel and Logistics

One hundred sixty-four unselected second year medical students and 25 tutors were involved. Two authors (JB & KA) gave two, two-hour, introductory lecture/demonstrations and issued a detailed handbook. The students were then divided randomly into groups of about six and each group allocated randomly to one of the 25 tutors with whom they stayed for six, two-hour sessions at weekly intervals. The 25 tutors were not unselected -- most were volunteers and some were specifically invited - but tutor selection was independent of the project. Three of the authors were tutors (J.B., S.C.-C,K.A.) The other tutors and all the students were blind to the project hypotheses.

Reprint requests should be addressed to Dr. Bird who is Visiting Professor in Liaison Psychiatry at the University of Alabama in Birmingham. Interview training at UAB has been run jointly by the Departments of Psychiatry and Internal Medicine for five years. Both departments supported this study.

The tutors comprised 7 psychiatrists, 9 internists, and 9 non-medical staff (7 psychologists and 2 social workers). Nineteen of the 25 attended one or more pre-course, two-hour, orientation session with the defaulters being evenly divided between groups. There were no gross differences between groups in average length of clinical experience or teaching experience. Of the four women, three were in the non-medical group.

Training content and procedures

In the handbook issued to all tutors and students, and in the introductory sessions, an explicit model of medical interviewing was presented which went beyond the traditional data gathering/instruction giving concept. The model was pragmatic (goal focussed) and largely behavioral. Interview goals and related skills were classified and illustrated under three heads - informational (history taking), emotional (helping feelings), and behavior change (compliance fostering). Content (history topics) detailed guidelines were also given. A model of the training process was also put forward. This can be summarized as orientation, demonstration, rehearsal with feedback, and practice. Tutors and students were shown how role-play training can usefully supplement work with actual patients. Outline session agendas were provided.

Instrumentation

A variety of measures of both process and outcome were used.

EDUCATIONAL PROCESS QUESTIONNAIRE FOR TUTORS (EPQ-T). This was presented to each tutor at the end of the course. For each major training topic (e.g. emotion skills) and each major training procedure (e.g. role-play) the tutor made three ratings on linear 1-5 scales to represent (a) past experience (b) actual time spent (c) value attached. All tutors responded.

STUDENT RESPONSE QUESTIONNAIRE (SR). Students were asked at the end of the course to rate aspects of the program. 81% completed the questionnaire. The key topics were:

How well was the process of medical interviewing presented?
(PRESENTATION OF INTERVIEW PROCESS) (PIP-SR)
How well was the assessment of interviewing presented?
(PRESENTATION OF INTERVIEW ASSESSMENT (PIA-SR)
How well was profiting from evaluation presented?
(PROFITING FROM EVALUATION) (PFE-SR)
Is there a need for further training?
WISH FOR FURTHER TRAINING (WFT-SR)
What is your overall impression of the course?
GLOBAL COURSE EVALUATION (GCE-SR)
How good are your own interview skills?
(INTERVIEW SKILL RATING BY STUDENT) (ISR-SR)

INTERVIEW SKILL RATING BY TUTOR (ISR-T). Tutors rated each of their students at the end of the course for information skills, emotion skills and behavior skills. ISR-T combines these ratings.

INTERVIEW SKILL RATING BY PATIENT (ISR-P) Every student recorded a 10 min. video-taped test interview just before their final tutorial. Sixteen volunteer out-patients from a cancer patient society were briefly trained to use two separate rating scales of interviewer skill: (a) clarity in handling facts and (b) rapport. Students were allocated at random and instructed to clarify facts of the present illness, and to show empathy and support. Patients recorded two scores:

INTERVIEW SKILL RATING BY PATIENTS-FACTS (ISR-PF)
INTERVIEW SKILL RATING BY PATIENTS-RAPPORT (ISR-PR)

KNOWLEDGE TEST (KT). All students answered 60 multiple choice questions at the end of the course. The first 35 questions concerned interview process and the second 25 concerned medical terminology and history topics. This test was constructed by composing a large pool of questions, having them answered by a panel of six experienced interviewers and then eliminating all that were not consistently answered by five out of six panel members.

Data Analysis

Data Reduction. The EPQ-T items were factor analyzed by the principal components method with iterations found in SPSS, Version 8.0 (Hall and Nie, 1979). Varimax rotations produced five procedure factors and four topic factors. (See Tables 1 and 2). Factor scores were computed using the approximate score procedure (Rummel, 1970) for each educational procedure and topic factor. These educational procedures and topic factor scores were then used to assess differences between tutor discipline groups.

Hypothesis Testing. One way analysis of variance was used since the appropriate unit of analysis is tutor means. Tukey's HSD post hoc test procedure (Hull and Nie, 1979) was used as the follow-up test. The statistical power of the one-way analysis of variance tests was assessed following Cohen (1977). At a significance level of 0.10 the power of the test is only 0.27. This makes the interpretation of non-significant findings ambiguous. As a result significance level probabilities as high as .10 will be considered adequate.

Results

A consistent trend is revealed in the means of the outcome variables given in Table 3. The non-medical group had the largest means for seven of the ten educational outcome variables. Table 4 reports the results of the one-way analysis of variance tests on the educational outcome variables. In addition to reporting the F-value and its significance level, Table 4 also reports eta-squared (η^2) and the approximate power of the test. Only one outcome variable, PFE-SR, reached a significance level of less than .05 and this favors the non-medical group.

Two outcome variables had a significant F-test at the .10 level - students rating of their interview skills (ISR-SR) and the patient's rating of the students rapport (ISR-PR). While post hoc tests revealed no one group that was significantly different from the others, examination of the means suggests that the difference between the internists and the psychiatrists accounted for the ISR-SR F-test. For the ISR-PR variable, the difference between the non-medical and psychiatrist group means probably accounted for the significant F-test. The eta-squared figures suggest that tutor's discipline is a relatively weak predictor of outcomes except for PFE-SR (self improvement training).

Regarding process variables the means and standard deviations are given in Table 5 and Table 6 displays the one-way analysis of variance, eta-squared and power. F values for role play and for video test emphasis are statistically significant at .10 or less. For role play only the non-medical group mean was significantly higher than the internist's using Tukey's HSD test. For video test emphasis, the difference between the non-medical group and internist's group means were beyond the .05 signifieance level using Tukey's HSD test. Eta-squared figures suggest that tutor discipline accounted for a moderate amount of variance in role play, while it accounted for much less in the remaining educational process variables.

Discussion

Interpretation clearly has to be cautious because the measures are of unknown reliability and validity and because the numbers are relatively small. Nevertheless, these non-medical tutors would seem to be at least as effective as medical tutors. This may be because they are clinicians working closely with doctors and may also reflect the fact that detailed guidelines were shared by all groups. It may be partly an artifact of loose measurement. However, it is a marked trend across various types of measure including a test of traditionally medical knowledge (history topics and medical terminology). Three product measures show significant differences between groups and two of these favor the non-medical tutors. Their students feel emphatically better equipped for self improvement and the patients they meet say they make better rapport. The trend in other measures supports the significance of these findings - non-medical tutors have the highest means on seven out of ten outcome measures.

Non-medical tutors would seem to differ from medical tutors in their approach to the training process. They are significantly different in their emphasis on role play and video testing with feedback, and tend to more emphasis on all forms of practice with feedback. Both the similarities and differences of approach may, paradoxically, be both accounted for by the explicit guidelines common to all groups - one might speculate that all groups were influenced by the guidelines but that the non-medical group followed them more closely. Whether the differences in training procedure account for the differences in outcome remains speculative. Certainly there is other evidence that role-play is a powerful training procedure for interpersonal skills. (Moreland et. al, 1976).

Conclusions

Psychologists and social workers using explicit guidelines can provide training in medical interviewing that is at least as good as that provided by doctors.

Acknowledgements

The authors are deeply indebted to Dr. Arthur Freeman III, M.D. for facilitating the project and contributing to the discussion of results. They also wish to acknowledge the important contributions of Dr. William Lukensmeyer in establishing this structure of medical interview teaching at the University of Alabama, of Ms. Carol Schaffhausen in being administrator to the course, of Dr. Nancy Hinson and other Office of Educational Development staff, and last but not least to Louis Josof and his patient colleagues in the TOUCH organization for providing invaluable feedback.

References

1. Barnes HV, Albanese M. Schroeder J, Reiter S. Senior medical students teaching the basic skills of history and physical examination. J Med Educ 1978; 53:432-434

2. Carroll JG, Monroe J. Teaching Clinical Interviewing in the health professions. A review of empirical research. Evaluations and the Health Professions 1980; 3:21-45

3. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977

4. Engel G. The need for a new medical Model:A challenge for bio-medicine. Science 1977;196:129-136

5. Helfer RE, Black MA, Helfer, ME Pediatric interviewing skills taught by non-physicians. Am J Dis Child 1975;129:1053-1057

6. Hull, CH, Nie NH SPSS Update. New York:McGraw-Hill, 1979

7. Moreland JR, Ivey AE, Phillips JS. An evaluation of microcounselling as an interviewer training tool. J Counsel Clin Psychol 1973;2:294-300

8. Rummel RJ Factor Analysis Evanston, IL:Northwestern Univ Press, 1970

9. Stillman PL, Sabers DL, Redfield BM. The use of paraprofessionals to teach intervieviewing skills. Pediatrics 1976;57:769-774

10. Stillman PL, Sabers DL, Redfield, BM Use of trained mothers to teach mothers to teach interviewing skills to first-year medical students: A follow-up study. Pediatrics 1977;60:165-169

Table 1

Educational Procedures Factors

VARIABLES	I[a]	II	III	IV	V
1. Experience with Feedback	.83				
2. Experience with Demonstrating	.69				
3. Experience with Student Practice	.68				
4. Experience with Actual Patients	.62				.52
5. Time in Demonstrations	.54			.54	
6. Experience with Video Tape	.45		.59		
7. Time in Role Play		.98			
8. Value of Role Play		.80			
9. Experience with Role Play		.53			
10. Time with Actual Patients		-.48			
11. Value of Video test			.82		
12. Time in Video test			.76		
13. Time in Feedback			.57		.53
14. Value in Feedback			.39		.55
15. Value of Practice with Actual Patients				.88	
16. Time in Practice with Actual Patients				.70	
17. Value of Demonstrations				.60	
18. Time in Student Practice					.39

[a] In order, the names of these factors are: Orthodox Training Experience (I Roleplay (II), Video Test Emphasis (III), Work with Actual Patients (IV), Practice and Feedback (V).

Tabel 2

Educational Topics Factors

Variable	I[a]	II	III	IV
1. Value of Medical Terminology	.84			
2. Value in History of Taking Skills	.80			
3. Time in Medical Terminology				
4. Time in History Taking Skills	.69	.36		
5. Experience in Problem Oriented Medical Records	.60	.60		.33
6. Time in Problem Oriented Madical Recrods	.54			
7. Time in Information Skills	.36	.78		
8. Time in Behavior Change Skills	.36			
9. Value of Information Skills	.33	.69		
10. Value of Behavior Change Skills	.32			
11. Experience with Medical Terminology		.73		.72
12. Experience in History Taking Skills		.59		
13. Time in Emotion Skills			.91	
14. Value in Emotion Skills			.83	
15. Value of Behavior Change Skilla			.65	
16. Time in Behavior Change Skills			.54	
17. Experience with Emotion Skills				.81
18. Experience with Information Skills				.74

[a] In order, the names of these factors are: Orthodox History Content (I), Orthodox History Process (II), Emotion and Behavior Change Skill (III), Clinical Experience (IV).

Table 3

Means and Standard Deviations of Educational Outcome Variables By Leader Discipline

Variables		Psychiatrists	Non-Medics	Internists
		Leader Discipline		
1. ISR-T	Mn	10.85	10.13	9.75
	SD	1.35	.64	1.16
2. ISR-SR	Mn	4.23	4.28	4.48
	SD	.31	.20	.18
3. ISR-PF	Mn	4.23	4.30	4.38
	SD	.15	.42	.44
4. ISR-PR	Mn	4.04	4.48	4.34
	SD	.32	.46	.33
5. KT-T	Mn	84.25	85.54	85.12
	SD	3.55	4.23	5.12
6. PIP-SR	Mn	3.70	4.10	3.89
	SD	.23	[illegible]	.34
7. PIA-SR	Mn	4.14	4.30	4.03
	SD	.22	.60	.35
8. PFE-SR	Mn	4.11	4.66	4.21
	SD	.25	.37	.21
9. GCE-SR	Mn	3.64	3.89	3.64
	SD	.27	.49	.42
10. WPT-SR		.62[a]	.69	.60

[a] Proportion of students indicating that they were interested in additional experience.

Table 4

ANOVA Summary Tables For Educational Outcome Variables

Variables	F[a]	P	η^2	P[b]
1. ISR-T	2.02	.16	.16	.36
2. ISR-SR	2.48	.10	.19	.38
3. ISR-PF	.29	.75	.02	.27
4. ISR-PR	2.56	.10	.21	.40
5. KT-T	.48	.63	.05	.21
6. PIP-SR	1.66	.22	.14	.34
7. PIA-SR	.84	.45	.08	.27
8. PFE-SR	8.06	.002	.45	.99
9. GCE-SR	.88	.43	.08	.27
10. WPT-SR	.71	.69	.08	.14

[a] All tests for violations of the homogeneity of variance assumption were not significant.

[b] P represents the power of the test given the obtained n^2 and a liberal alpha probability of .10.

Table 5

Means and Standard Deviations of Educational Process Variables By Leader Discipline

Variable		Psychiatrists	Non-Medics	Internists
		Leader Discipline		
Procedures				
Orthodox Training Experience	Mn	-.52	.91	-.37
	SD	2.89	.94	1.84
Role Play	Mn	.1[illegible]	1.28	-1.42
	SD	1.54	.69	1.51
Video Test Emphasis	Mn	-.01	.38	-1.31
	SD	1.47	1.35	1.21
Work with Actual Patients	Mn	.23	.07	.08
	SD	1.15	1.39	1.27
Practice with Feedback	Mn	-.39	.40	.37
	SD	1.81	.85	1.05
Topics				
Orthodox History Content	Mn	-.39	.09	.78
	SD	3.62	2.97	2.29
Orthodox History Process	Mn	-.01	.32	.61
	SD	1.29	1.89	1.93
Emotional and Behavior Change Skills	Mn	.27	.14	.95
	SD	1.27	1.75	2.10
Clinical Experience	Mn	-.12	1.16	-.58
	SD	2.22	.79	2.16

Table 6

ANOVA Summary Tables For Educational Process Variables

Variable	F[a]	P	η^2	P[b]
Procedures				
1. Orthodox Training Experience	1.13	.35	.11	.40
2. Role Play	8.81	.001	.47	.99
3. Video Test Emphasis	3.15	.06	.25	.78
4. Work With Actual Patients	.04	.96	.004	7.05
5. Practice With Feedback	.95	.41	.09	.34
Topics				
6. Orthodox History Content	.25	.78	.04	.21
7. Orthodox History Process	.46	.64	.05	.22
8. Emotion and Behavior Change Skill	1.07	.36	.10	.27
9. Clinical Experience	1.39	.28	.15	.41

[a] All tests for violations of the homogeneity of variance assumption were not significant

[b] P represents the power of the test given the obtained n^2 and a liberal alpha probability of .10

PRECIS

TEACHING IN THE CLINICAL SETTING

MODERATOR: John Littlefield, Ph.D.
University of Texas Health Sciences Center

ANALYSIS OF CLINICAL EXPERIENCE

Little is known about the clinical problems encountered and procedures performed by medical students during their predoctoral training. A computerized program to document the total clinical content of the medical curriculum is described, the advantages and methodology outlined, and preliminary results from the first year summarized.

IDEAL AND ACTUAL RESIDENT TEACHING PRACTICES IN A UNIVERSITY HOSPITAL

A study was conducted concerning desirable and actual teaching practices as perceived by residents in a university hospital. The method identified areas in which actual practice was perceived to coincide with what was felt to be desirable and areas in which this was not the case thus allowing priorities for change to be established.

SIMILARITIES OF GENERAL MEDICINE CLINIC IN A TEACHING HOSPITAL TO INTERNAL MEDICINE PRACTICE

This study was undertaken to determine if a general medicine clinic in a teaching hospital provided an experience similar in content to that seen in the office of a general internist. Data on 4856 visits to a university clinic in 1979 were collected on encounter forms and were compared to data on office visits to internists published by the National Ambulatory Medical Care Survey.

THE RELATIONSHIP BETWEEN MEDICAL STUDENT CLERKSHIP ACTIVITIES AND PERFORMANCE ON NBME PART II

Student-completed logs were used to collect information regarding student activities in required Internal Medicine clerkships at four different hospitals. While significant differences in experiences existed among the hospitals, no differences in student performance on the Internal Medicine section, NBME Part II, were detected based upon the amount of time students devoted to patient care or educational experiences.

ANALYSIS OF CLINICAL EXPERIENCE --
A Preliminary Report

by
Collin Baker, M.D.
Department of Family Medicine
University of South Carolina School of Medicine
Columbia, S.C.

Of the several ways in which medical education may be evaluated, most medical schools employ evaluation of <u>outcome</u>, measured by course tests and the ability of their graduates to successfully pass Parts I and II of the National Boards. Evaluation of the <u>process</u> of education is less often used, since it is more difficult to carry out, especially in the clinical areas.

Until now there has been only rudimentary documentation of the clinical content of medical education, in contrast to the basic science content. In 1966, one student, Keith Hodgkin, documented his total clinical experience in medical school and internship, comparing it with the problems he encountered in actual practice.[1] Garrard and Verby published a study in 1977 in which they compared data from the clinical experiences of a limited number of medical students in a rural training program with those of a control group in the clinical rotations at the University of Minnesota Medical School at Minneapolis.[2]

More recently, Parkerson and colleagues published analyses of clinical problems encountered by medical students in rural settings as compared with those in the Duke University Medical Center.[3,4] Other fragmentary and departmental studies are found,[5,6,7,8] but the data thus recorded is limited; only Hodgkin and one of the Parkerson studies (6) recorded the total clinical experiences of students and only three students were thus involved.

If clinical competence and preparedness for postgraduate training are at least in part determined by the scope of undergraduate clinical training, it would seem important to examine the content of clinical experience in order to ascertain whether the training offered students is adequate. In recent years, there has been increasing concern about the clinical skills of first year house officers. Documentation of the kind and number of clinical procedural skills practiced during the undergraduate years should prove of value to students applying for residency training by documenting their level of proficiency.

Such a method of evaluation should be of benefit not only to the student, but to the school and to the individual departments. Students could use the information to direct their choice of electives to fill gaps in their training. Departments could examine the training given in different settings and under different instructors as a guide to improving teaching methods. Curriculum Committees could base curricula on concrete data, making modifications as necessary to provide a comprehensive and well-rounded medical education.

*Reprint Requests: Collin Baker, M.D., Department of Family Medicine, University of South Carolina, 3301 Harden Street, Columbia, SC 29203.

Method

A program of this kind has been initiated at the University of South Carolina School of Medicine. This is a developing school in which clinical teaching is done in three affiliated institutions -- Richland (County) Memorial Hospital, the Veterans Administration Hospital at Columbia and the William S. Hall Psychiatric Institute. Students on clinical clerkships, beginning in their third year, use a problem-oriented record system which is uniform for all affiliated institutions. The problem list of the student record incorporates a data recording system which is based on the ICHPPC diagnostic index.[10] Since this coding system was designed for use chiefly in primary care, selected rubrics have been added to provide codes for problems common to the other specialties; all code numbers are compatible with the ICD-9-CM.[11]

The same sheet also serves for recording clinical procedures performed by the student on each service rotation, and for indicating the degree of student involvement. Data from the outpatient experience are recorded on a pocket-size slip; data thus recorded are summarized in Table I. These Record of Clinical Experience (RCE) forms are collected at the end of each clinical rotation and are then forwarded to a central office, where they are coded by a code clerk and entered into a computer. Coding by a single clerk, rather than by each student, assures uniformity of coding and results in more reliable data. Quarterly printouts are provided to the students and departments. Summary sheets are prepared for the Office of the Dean for Academic Affairs and the Curriculum Committee.

Results

Abridged data from the first year of operation of the program are shown in Table I, which lists problems and procedures in order of frequency. The sorting by departments allows determination of which problems are most frequently encountered in each department.

Table II is a partial departmental summary, abbreviated to the ten most common problems seen in that department. The degree of involvement of students in each problem and procedure is indicated.

The student is chiefly interested in accumulating data that show the scope of his experience as he progresses through clinical training. For this reason, each student receives a quarterly summary which lists the number of encounters for each problem and procedure and the degree of his involvement. Table III abstracts such a summary.

Discussion

Although we have not yet been able to provide students or faculty with complete data, acceptance of the program has been remarkably good. All but two of the students submitted data for tabulation in the initial months, and at least *2/3 faithfully continued recording their data throughout the year. There has been enough interest among the departments that the data has been collected regularly, and requests for certain of the data have already been received from the Curriculum Committee.

Startup costs to date have been about $2500, including initial analysis, program design, and implementation costs. Computer time, coding costs, and storage charges are now estimated at less than $30 per student per year. Exact figures will be available when the year's tabulations have been completed.

Conclusions and Implications

A method of recording data from clinical experience of medical students has been described, and initial data from one school presented. While the data promises to be of considerable value within the institution initiating the program, it should be of even greater value if other schools adopt similar programs. This would allow comparison of the process of clinical training in different clinical settings, in different parts of the country, and under different curriculum plans, and might thus improve the quality of education of medical students throughout the country.

*To be updated when data is complete for the year.

TABLE I

[illegible]ARTMENTAL SUMMARY

AVERAGE N[illegible] OF PROBLEMS PER STUDENT

PROBLEM	FAMILY MEDICINE	INTERNAL MEDICINE	NEURO PSYCHIATRY	PEDIATRICS	SURGERY	OB-GYN	TOTAL
DIABETES MELLITUS	21.75	13.20		1.50	0.60	1.80	38.85
HYPERTENSION, UNCOMPLICATED	25.11	10.20		0.75		0.60	36.72
PREGNANCY, UTERINE						16.20	16.20
DEPRESSIVE DISORDERS	9.75		4.50			0.60	14.85
PAROXYSMAL TACHYCARDIA	4.86	9.00					13.86
CONGENITAL ABNORMALITIES, NEC	0.75	0.60		11.25			12.60
OTHER HEART DISEASE, NEC	1.86	4.20	0.60	0.75	4.20		11.61
PNEUMONIA	2.25	5.40		3.00	0.60		11.25
HEART MURMUR, NOT YET DIAGNOSED				11.25			11.25
OBESITY, DIETARY	10.11	0.60			0.20		10.91
EMPHYSEMA AND C.O.P.D.	2.61	6.60	0.36		0.60		9.66
CONVULSIONS	1.86	3.60		3.00			8.46
SCHIZOPHRENIA	1.50	0.60	6.36				8.46
URINARY TRACT INFECTION	5.61	1.20		1.50			8.31
OTHER DIS. RESPIRATORY SYSTEM	1.11	4.20			1.80		7.11
GOUT AND HYPERURICEMIA	5.25	1.80					7.05
OTHER DISEASE OF NERVOUS SYSTEM				6.75			6.75
FEVER OF UNDETERMINED ORIGIN	1.11	2.40		3.00			6.51
CIRRHOSIS AND OTHER LIVER DISEASE	2.25	4.20				0.60	6.45
ANXIETY STATES	5.25	1.11					6.36
SEROUS OTITIS MEDIA				6.00			6.00
OTHER GENITOURINARY DISORDER	3.75	1.80					5.55
ACUTE OTITIS MEDIA				5.25			5.25
CONGENITAL ABNORMALITY OF HEART				5.25			5.25
ELEVATED BLOOD PRESSURE, NYD	5.25						5.25

PROCEDURES	FAMILY MEDICINE	INTERNAL MEDICINE	NEURO PSYCHIATRY	PEDIATRICS	SURGERY	OB-GYN	TOTAL
DELIVERY, SPONT VAG						12.40	12.40
LUMBAR PUNCTURE	1.80	4.20		1.50			7.50
CAESAREAN SECTION						6.40	6.40
IUD INSERTION						4.20	4.20
THORACENTESIS	0.60	2.40					3.00
[illegible]CHOSCOPY					1.80		1.80

TABLE II
DEPARTMENTAL SUMMARY

FAMILY MEDICINE — RMH OUTPATIENT

PROBLEMS	PRIMARY RESPONS	SHARED RESPONS	EXAMINED ONLY	OBSERVED ONLY	TOTAL	AVG PER STUDENT
HYPERTENSION, UNCOMPL	96	0	96	3	195	24.36
DIABETES MELLITUS	51	0	54	9	114	14.25
OBESITY, DIETARY	60	0	15	6	81	10.11
DEPRESSIVE DISORDERS	48	0	18	0	66	8.25
ANXIETY STATES	9	0	12	3	42	5.25
ELEVATED BLOOD PRESSURE NYD	36	0	3	0	39	4.86
URINARY TRACT INFECTION	21	0	12	3	39	4.86

PROCEDURES	PERFORMED $\bar{S}$ SUPERV	PERFORMED $\bar{C}$ SUPERV	ASSISTED	OBSERVED ONLY	TOTAL	AVG PER STUDENT
URINALYSIS	43	8	0	0	51	6.38
EXCISION SKIN LESION	0	34	1	0	35	4.38
SIGMOIDOSCOPY	0	22	0	4	26	3.25
LUMBAR PUNCTURE	0	3	0	1	4	0.50

(NOTE: COVERS ONE ROTATION ONLY)

TABLE III
STUDENT SUMMARY

STUDENT: W________ C________

INTERNAL MEDICINE — VAH INPATIENT

PROBLEMS	PRIMARY RESPONS	PRIMARY RESPONS	EXAMINED ONLY	OBSERVED ONLY	TOTAL
DIABETES MELLITUS	6	0	3	2	11
HYPERTENSION, UNCOMPL	5	0	2	0	7
PAROXYSMAL TACHY	2	0	3	1	6
EMPHYSEMA AND COPD	0	0	4	2	6
ASCITES	2	0	3	1	6
ALCOHOL ABUSE	0	0	3	2	5
CONVULSIONS	0	0	4	1	5

PROCEDURES	PERFORMED $\bar{S}$ SUPERV	PERFORMED $\bar{C}$ SUPERV	ASSISTED	OBSERVED ONLY	TOTAL
BRONCHOSCOPY	0	0	0	1	1
LUMBAR PUNCTURE	0	2	0	0	2

REFERENCES

1. Hodgkin, K. 'Toward Earlier Diagnosis', Livingstone, Ltd., London, 1966.

2. Garrard, J., and Verby, J E, 'Comparisons of Medical Student Experiences in Rural and University Settings' J. Med. Ed. 52:802, Oct 77.

3. Senf, J H, and Vuturo, A F, 'Differences in the Clinical Clerkship Experience of Female and Male Students' J. Med. Ed. 53:517, June 78.

4. Friedman, C P, Stritter, F T, and Talbert, L M, 'A Systematic Comparison of Teaching Hospital and Remote-Site Clinical Education' J. Med. Ed. 53:565, July 78.

5. Parkerson, G R, and Baker, C, 'Clinical Experience of Medical Students in North Carolina Family Medicine Preceptorships' J. Med. Ed. 55:42, Jan 80.

6. Parkerson, G R, Bencze, R F, Griffin, E W, and Baker, C, 'Clinical Experience of Medical Students with Primary Care Career Goals' J. Fam. Prac. 9:639, 1979.

7. Yager, J, 'The Clinical Experience Worksheet' J. Med. Ed. 50:1061, Nov 75.

8. Kaufman, F, Khan, M, 'A Comparison in Inpatient Clinical Experience Between a University and Two Community Hospital Pediatric Services', Md. State Med. Journal 23:46-47, May, 1974.

9. Holden, W D, 'The Interface Between Undergraduate and Graduate Medical Education', JAMA 241:1148-1150, March, 1979.

10. 'International Classification of Health Problems in Primary Care' (ICHPPC-2). An adaptation of the International Classification of Diseases (9th revision), prepared by the Classification Committee of the World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians (WONCA) in collaboration with the World Health Organization.

11. White, K L, (Chairman), International Classification of Diseases, 9th revision, Clinical Modification, (Vols 1 & 2), published by the Committee on Professional and Hospital Activities of the U.S. National Committee on Vital and Health Statistics, 1978.

Ideal and Actual Resident Teaching Practices in a University Hospital.

Carter Zeleznik, Ph. D., and Paul Brucker, M.D., Jefferson Medical College, Philadelphia, Pennsylvania

Introduction. Much of the clinical teaching of medical students in the United States is done by residents. Teaching also represents a major activity of residents. (1) Since the transition from the status of learner to that of teacher takes place shortly after graduation from medical school in many cases, it is reasonable to inquire as to how well prepared residents are for teaching. Except in rare situations, there is no formal instruction in teaching given to residents. (2) Suggestions have been made about how to improve the teaching of residents but it is not clear that these have been actued upon. (3)

If it is conceded that how medical students are taught and how they learn during their clinical clerkships is of importance, it follows that deficiencies identified in the teaching practices of those who provide such educational experience have the potential for becoming self-replicating because of the short time interval separating learner status from student status. This brief period of time thus makes change within the system exceedingly difficult. Perhaps for this reason, as judged by the paucity of literature on the subject, medical educators have avoided examination of the matter. (2) Whatever the present complaints, they may be less discomforting than collecting data which would make it still more difficult for the problem to be dismissed.

Nevertheless, as part of the self-flagellation that medical schools participate in at periodic intervals, the Curriculum Committee at Jefferson Medical College requested that a study be made concerning the residents' role in teaching undergraduate medical students. The purpose of this paper is to present information about the methodology followed in performing this study and in presenting resulting data to the institution's Curriculum Committee.

Methods. On the basis of anecdotal data concerning residents' teaching activities at the university hospital and published information from other institutions, a set of 47 propositions was formulated describing desirable characteristics or criteria of a residency teaching program. These were, in turn, categorized in four main areas: 1) preparation of the residents for teaching, 2) methods of instruction used by the residents, 3) techniques of evaluation of residents' teaching, and 4) rewards associated with residents' teaching. Specific hypotheses were generated as to the likelihood of actual practice conforming to the desired criteria. In general, the hypotheses were pessimistic, based again upon

Send reprint requests to: Carter Zeleznik, Ph. D., Associate Director, Office of Medical Education, Jefferson Medical College, 1025 Walnut Street, Philadelphia, Pennsylvania, 19107.

anecdotal reports and information available in the literature.

The propositions generated were then used in the creation of a questionnaire in which residents at the university hospital were asked to indicate the extent to which a given characteristic was desirable. Some propositions were stated in the negative so as to encourage avoidance of unthinking agreement with the entire set of statements as desirable. Residents were also requested to indicate whether or not the specific criterion represented in the statement conformed to actual practice in the residents' respective departments. (Copies of the complete questionnaire are available upon request.) Some examples of items are as follows:

Preparation: Residents have the option of whether or not to teach undergraduate medical students.

Methods of Instruction: Residents provide clear-cut educational objectives to medical students they are teaching.

Techniques of Evaluation: Residents' evaluations of medical students are reviewed by members of the attending faculty.

Rewards: Residents are given faculty appointments in the medical school when they become senior residents.

With regard to the first example given, it was expected that residents would agree to the desirability of having an option about whether or not to undertake teaching responsibilities. However, it was also expected that such responsibilities would be assigned to them whether or not they wished to have them (actual practice).

Questionnaires were distributed to 245 residents in eight departments at the university hospital with the permission of the chairmen of the departments who had reviewed the questionnaire and had endorsed it. Residents were encouraged to sign their names on the answer sheets but in any event to identify their departmental affiliation. A total of 111 responses were received. The low response rate was expected and no doubt reflects the relatively low priority given to teaching in some departments and by many residents. It may also reflect the difficulty of locating some of the residents. Questionnaires were mailed three times and department chairmen urged the residents to complete them.

Given the responses of the residents, tabulations were made for presentation to members of the curriculum committee. However, prior to providing them with this information, each individual was asked to complete the questionnaire indicating what he or she felt to be desirable. Upon completion of this exercise, tabulations were opportunity to select among research, service, and teaching for emphasis in their individual programs (66% to 11%) but nearly hafl of the members of the Curriculum Committee were uncertain or unwilling to accept this as being desirable. As might be expected, the majority (64%) of residents said this was not a prevalent practice in their departments and only 15% said that it was.

With regard to individual items, correlations ranged between .68 and -.11 in terms of what residents described as desirable and what they said was the actual practice in their departments. Correlations below .20 are not significant at the .05 level of probability given the number of respondents. Nevertheless, thirty three of the forty seven correlations were significant at or beyond this level. This suggests that in spite of lack of complete conforimity between desired practice and actual practice, there is a positive relation between the two.

Rank order correlations comparing desirable educational practice as described by the residents and as described by members of the Curriculum Committee was surprizingly high, .76 (p = .001). A somewhat lower rank order correlation was computed comparing what the residents saw as desirable overall and what they perceived actually to be the case .46 (p = .01). A still lower rank order correlation was found between Curriculum Committee statements as to what was desirable and what residents stated was the actual practice in their departments, .28 (marginally significant, p = .06).

Discussion: The level of consensus as indicated by the high correlations as to what is desirable in residents' teaching between members of the Curriculum Committee at Jefferson Medical College and residents in training at its hospital indicates that people do not need to be told what is or is not desirable concerning these matters in general. While some disagreement exists, a large number of individuals at this institution, as judged by responses to a questionnaire, appear to agree with points of view discussed in the literature on the subject. It may even be observed that there is a positive relation between what residents perceive as desirable and what they describe as actually the case both in the aggregate and item by item.

Examination of specific items with high and low correlations provides indicators of areas of particular strength and weakness in the residents' teaching activities at this institution. Some of these may require institutional action whereas others may be corrected on a departmental basis. Because the curriculum committee has representation from each of the departments on it, communication concerning identified areas where desirable practice is not perceived as actual practice is facilitated. Moreover, both positive and negative aspects of departmental educational programs may be recognized.

ade of the committee's responses and data from the residents and from the committee were distributed to the committee.

Responses to each item were scored on a five point Likert-type scale with the following weights: Very desirable practice (D) = +2, Desirable practice (d) = +1, Uncertain whether or not desirable practice (?) = 0, Undesirable practice (u) = -1, and Very undesirable practice (U) = -2. A similar code was established for scoring whether or not the practice was seen in the resident's department. A weight of 0 was also assigned to omitted items.

Total scores for each item were tabulated for each of the two score categories mentioned. Scores were computed separately for residents and for members of the curriculum committee. Pearson product moment correlations were computed for the two dimensions of the scale for responses given by residents. Spearman rank order correlations were computed comparing responses between the two groups.

Results. There was a high level of consensus between the two groups with regard to many of the items as reflected in terms of total scores for what was considered desirable. For example, 67% of the residents expressed the belief that teaching responsibilities should be assigned to residents contingent upon their demonstrating skill in teaching or having received instruction in doing so. 22% were uncertain and only 8% indicated that it was undesirable. The remaining 3% did not respond to the item. Only one member of the Curriculum Committee felt that this was undesirable and 15 members felt that it was desirable. However, less than 5% of the residents reported that this was a customary practice in their departments and 83 percent reported that it was not a customary practice. Eleven percent were not certain about this in their departments.

There was agreement in both groups that residents should be informed when applying for positions on the house staff that they would be expected to teach undergraduate students. Rather unexpectedly, it was found that 64% reported this to be an actual practice in their departments and only 20% indicated that it was not.

Residents also felt that they should discuss the evaluations they make of medical students with each student individually, 70% finding the criterion desirable and only 8% finding it undesirable. A strong majority of the curriculum committee concurred in this (13 to 3). However the data indicate that this is rarely done in actual practice (15% to 59%). Similarly, both residents and Curriculum Committee members agreed that it was desirable for residents and departmental faculty to meet regularly to discuss educational matters (86% to 3% for the residents and 16 to 0 for the Curriculum Committee) but only 7% of the residents reported that this was an actual practice in their departments and 88% reported that it was not. On the other hand, residents felt it desirable for them to have the

Thus the data indicate that residents and members of the Curriculum Committee believe that residents should delegate clinical responsibilities to students only when the students have demonstrated clinical competence in those specific areas. The data also indicate that the residents perceive such to be the case in most instances already. On the other hand, the data indicate that although residents believe that their program directors should review evaluations of their teaching made by students, this is not actually done. In this instance, corrective action may be taken relatively easily.

In some regards, the situation insofar as it pertains to residency teaching practices is not greatly different from what Brown described in Virginia several years ago. Our data are also consistent with what Tonesk has described as prevalent in many medical schools with regard to lack of formal assessment of residents' teaching. The problem remains to increase the attention given to evaluating residents as teachers and of providing them with instruction in areas of weakness. If one were to attempt to implement large scale training programs for residents to learn how to be better teachers, extensive reprioritization in the use of institutional or departmental resources would be required. On the other hand, by looking at specific practices, developing consensus as to what is important and what is less important and what is and what is not presently the case, the possibility of addressing specific issues and producing change incrementally is raised. While it may not be feasible for institutions or departments to provide residents with extensive instruction in educational methodology, other issues may be more conveniently addressed. Review of items with low correlations between what is perceived as desirable and what is perceived as actually the case reveals, for example, that residents and departmental faculty do not meet regularly to discuss educational objectives, the progress of students, or educational problems which have arisen even though there is a general consensus as to the desirability of this occurring. Change in this and other areas may not be so difficult.

<u>Conclusion</u>: A technique has been described for determining the level of agreement of different groups of individuals with regard to the desirability of certain practices associated with residents' teaching in a university hospital. The technique also permits determination of the extent to which actual practice is perceived to conform to desired practice: On the basis of data obtained using the method presented, priorities may be established for corrective action.

Initial data suggest that there is a higher overall conformity between what residents believe to be desirable and what they see as actually the case than between what members of the institution's Curriculum

Committee believe to be desirable and what the residents see as the actual situation. By confronting the Curriculum Committee and through it departmental chairmen with such data, it is expected that selective change may occur. The method also provides a means by which change, if it occurs, may be monitored over time.

References:

1. Stefanu, Constantine, Pate, Mary L., and T. Albert Farmer, Jr., A Management Model to Survey and Analyze House Staff Effort, JME, 51: 50-52, 1976.

2. Brown, Robert S., House Staff Attidues Toward Teaching, JME, 45:156-159, 1969.

3. Tonesk, Xenia, The House Officer as a Teacher: What Schools Expect and Measure, JME, 54: 613-616, 1979.

Similarities of General Medicine Clinic in a Teaching Hospital to Internal Medicine Practice

Roberta A. Monsón, M.D. and Judith Jameson, M.P.H.
University of Arkansas Medical Sciences Campus

During the past decade there has been a notable trend in including more ambulatory care experience as part of internal medicine residencies - both traditional programs and primary care residencies. As the emphasis on ambulatory care in academic medical centers increases, the question of whether teaching hospitals provide a good setting for such training has been raised. One way of examining the potential value of such an experience is to review the types of patients seen in a university clinic, their problems and diagnoses, and to determine how closely these patients would resemble those seen in a general internist's private practice.

One study of a teaching hospital population has shown that the prevalence of certain diagnoses is significantly higher than one would expect in a standard office practice and that several uncommon medical problems were seen within a sample of 271 patients.[1] As part of our resident evaluation program, information is collected on visits to a general medicine clinic in a university hospital.[2] We have analyzed this data to determine if the population seen in our clinic resembles that seen by internists in office based practice extensively described by the National Ambulatory Medical Care Survey (NAMCS).[3,4]

Methodology

Data on all patient visits to the University of Arkansas General Medicine Clinic during 1979 were collected on encounter forms at the time of the visit and batch processed by computer. Providers of care included faculty general internists, medicine residents, fourth year medical students and nurses. Providers indicated the prior visit status of the patient, duration of visit, services ordered (laboratory or x-ray), referrals, disposition, problems or diagnoses dealt with on that visit and drugs prescribed and entered their identification number. A report was generated to describe the above parameters for the period 1/1/79 through 12/31/79. Results of laboratory tests are not described in this study because each test (e.g. sodium or potassium) was ordered as a separate unit rather than as a panel. Data from this report were analyzed and compared with data from the NAMCS.

Patients seen in the university medicine clinic come from several sources; they may be seen as a follow-up for hospitalization or an emergency room visit, referred from other practicing physicians, referred from other university clinics or self-referred. No routine laboratory studies are pre-ordered so that all ancillary service use reflects decisions by the provider.

Reprint Requests: Roberta A. Monson, M.D.
University of Arkansas Medical Sciences Campus
4301 W. Markham, Slot 555A
Little Rock, AR 72205

Patients seen by housestaff and students are presented to faculty attendings; eighty percent of the faculty attending time is provided by general internists.

Demographic data are compared to NAMCS data, though age groupings differ slightly as noted. A sample of 2500 active patients were reviewed for racial distribution.

Referrals from the university setting include not only referrals to other physicians, clinics and agencies but also to social service, dietetics, and mental health services and were often given with a return appointment to the provider.

Results

Analysis of 4856 clinic visits to the university compared to a sample of 8599 of the projected 62,117,000 visits to internists described by the NAMCS are presented. As seen in Table 1, the 15 most frequent diagnoses made by NAMCS physicians comprise 45% of the visits while 12 of those 15 diagnoses were made on 49.2% of visits to a general medicine clinic. Significant differences in the prevalence of hypertension, diabetes and obesity are obvious and probably reflect the population being studied. The average number of problems dealt with per encounter was 1.47. Demographic data are shown in Table 2. 58% of medicine clinic visits were made by whites, 42% by non-whites. Ninety-one percent of NAMCS visits were made by whites and 9 percent by non-whites.

Table 3 shows the duration of physician-patient encounters. As can be seen, 91% of the visits to NAMCS physicians were completed within 30 minutes while only 75% of visits to a university physician were completed within this time. Radiologic studies were ordered by NAMCS physicians on 13% of visits while 16.4% of clinic visits included radiologic procedures. In contrast, NAMCS physicians obtained electrocardiograms on 14% of their patient encounters versus 8.7% in the university setting.

Disposition of patients is shown in Table 4. There are significantly more ($p < 0.001$) admissions and referrals in the university setting than in NAMCS offices. Specific return appointments are given significantly ($p < 0.02$) more often in the physicians private offices than in the clinic setting. The most frequent referral sites in the university setting were dietetics, dermatology, ophthalmology and ENT.

Discussion

With the increasing use of teaching hospitals to train physicians for primary care it is appropriate to examine the type of patient problems seen in such a setting and to establish their similarity or difference to problems seen by the practicing physician. The extensive data compiled by the NAMCS give us a unique basis for identifying the experience of the practicing physician. In the present study, we have systematically collected data on patients attending a general medicine clinic in a teaching hospital of the state's only medical school. While this might tend to bias the experience toward more "secondary" care, we found that the problems encountered most frequently were those also most frequently seen by physicians in the NAMCS. Perhaps the most notable exception was upper respiratory infections which ranked fifth in frequency in

NAMCS and was not included in the top 25 problems seen in the university clinic. Certain problems like hypertension, obesity and diabetes were more prevalent in our population than in NAMCS which may reflect the different demography of our population.

Relatively few studies describe teaching hospital ambulatory populations in detail. Fletcher et al,[1] describe such a population in an urban teaching hospital where general medical and medical subspecialty patients are seen in a consolidated "medical polyclinic." Subspecialists were involved in 64 percent of the visits and general internists in 36 percent. Not surprisingly they report a disease prevalence significantly different from our population and that of the NAMCS. Further, patients seen in their clinic averaged over 2.7 problems per patient visit compared to 1.5 problems/visit, in our study suggesting that the population using their ambulatory services were more medically complex.

The results of this study suggest that our patients have problems similar to those seen in the NAMCS and that few problems seen by NAMCS are not seen in our population even though the prevalance or some problems may differ significantly. It is likely that similar studies in other teaching institutions may also show considerable variation in the prevalence and variety of illness seen in ambulatory populations. This variation may well reflect geographic and demographic variations as well as the subspecialty versus general clinic design. What is most important is that housestaff be aware of the similarity or difference in prevalence from their clinic population to that in an office based practice. In doing so, they can be taught the likelihood of identifying the more uncommon problems and of applying cost-benefit principles to the selection of patients for more extensive diagnostic workups. The NAMCS data demonstrate that internists frequently order laboratory and radiogic studies, particularly when compared to family practitioners.[3] This may reflect the emphasis on inpatient, diagnosis-oriented care which has prevailed in training programs in decades past. Perhaps the recent addition of more ambulatory care experience where "common things are common" and the current efforts at cost containment will influence practice habits of future practitioners.

REFERENCES

1. Fletcher SW, Fletcher RH, Pappius EM, Rudd, R. A teaching hospital medical clinic: secondary rather than primary care. J Med Educ, 1979, 54:384-391.

2. Monson, RM, Jameson, J. An Ambulatory Care Information System for Evaluating Clinic Performance of Medicine Residents. Proceedings of Seventeenth Annual Conference on Research in Medical Education. Assoc Amer Med Colleges. 171, 1978.

3. National Center for Health Statistics. The National Ambulatory Medical Care Survey: 1975 summary. Rockville, Md. The National Center for Health Statistics, 1978, (Vital and health statistics. Series 13: no. 33) (DHEW publication no. (PHS) 78-1784).

4. Idem. National Center for Health Statistics. The National Ambulatory Medical Care Survey: Office visits to Internists. Rockville, Md. The National Center for Health Statistics, 1979. (Vital and health statistics. Series 13: no 36) (DHEW publication no. (PHS) 79-1787).

5. Noren J, Frazier T, Altman I, DeLozier J. Ambulatory medical care: A comparison of Internists and Family-General Practitioners. N Engl J Med. 1980; 302:11-16.

Table 1

Percent of Visits by Problem/Diagnosis in NAMCS and University Medicine Clinic

NAMCS Rank	Problem/Diagnosis	NAMCS % Visits	Medicine % Visits
1.	Essential hypertension	9.3	18
2.	Chronic ischemic heart disease	7.9	3.7
3.	Diabetes mellitus	4.5	8.9
4.	Medical or special examination	4.1	--
5.	Acute upper respiratory infection	2.6	0.5
6.	Neuroses	2.3	2.5
7.	Osteoarthritis and allied conditions	2.3	2.5
8.	Symptomatic heart disease	2.0	4.0
9.	Medical and surgical aftercare	1.8	--
10.	Rheumatoid arthritis and allied conditions	1.6	0.5
11.	Obesity	1.6	5.8
12.	Observation without need for further medical care	1.3	--
13.	Emphysema	1.3	1.9
14.	Hay fever	1.2	.34
15.	Other eczema and dermatitis	1.2	1.4
	% Total Visits	45.0	49.2
	Total Visits Reviewed	8599	4856

115

TABLE 2
DEMOGRAPHIC DATA
University Clinic Compared to NAMCS
January 1, 1979 - December 31, 1979

FEMALE

Clinic Age	Clinic %Visits	NAMCS Age	NAMCS %Visits
0-19	1.14	0-15	2.7
20-29	6.15	16-24	8.9
30-39	8.04	25-44	20.1
40-49	9.78	45-64	36.7
50-59	15.61	65-over	31.6
60-69	12.06		
70-79	7.36		
80-over	1.94		

MALE

Clinic Age	Clinic %Visits	NAMCS Age	NAMCS %Visits
0-19	.51	0-15	4.2
20-29	3.75	16-24	8.7
30-39	3.75	25-44	22.5
40-49	3.82	45-64	39.8
50-59	5.52	65-over	24.8
60-69	6.69		
70-79	6.57		
80-over	1.89		

TABLE 3
DURATION OF VISIT

Minutes/Visit	Clinic	NAMCS
0 min	0.7	0.4
1-15 min	66.0	40.1
16-30 min	24.6	34.2
31-60 min	7.8	14.0
61-or more	0.9	4.0
unknown	7.5	---

TABLE 4
DISPOSITION OF VISIT

Disposition	% of Visits*	
	Clinic	NAMCS
no return or prn	15.6	25.6
return visit	56.2	68.4
referral to another provider	9.6	4.4
hospital admission	3.73	1.7

*Percents will not add up to 100 because some patient visits required more than one disposition

117

THE RELATIONSHIP BETWEEN MEDICAL STUDENT CLERKSHIP ACTIVITIES AND PERFORMANCE ON NBME PART II

Judith G. Calhoun, Ph.D., Alan L. Hull, Ph.D.
and Wayne K. Davis, Ph.D.
The University of Michigan Medical School*

Introduction and Objective

During the past decade an increasing proportion of clinical medical education has been transferred from medical center hospitals to governmental and community hospitals affiliated with medical schools. General concern has been expressed regarding the lack of curricular standardization and the resultant outcomes due to this distribution of educational experiences to differing sites.[1,2] Although several studies have failed to substantiate the relationship between student performance on standardized tests and the site of the clerkship experience, the bias remains in favor of the classic medical school teaching hospital and its assumed educational superiority as compared to its affiliates. Equally as persistant are the stereotypes that exist regarding these educational settings and the types of clerkship experiences that they provide. These characterizations usually depict the teaching hospital as the citadel of learning with more time being spent on secondary and tertiary patient care and educational activities such as lectures, rounds and conferences. In contrast, the community or private hospital affiliate is viewed as emphasizing primary care and experiencial activities and frequently allowing students greater flexibility and responsibility. Governmental affiliates such as Veteran's Administration hospitals and county hospitals are also thought to offer students more active roles in patient care but with a different type of patient population; predominately the indigent with chronic and far-advanced disorders.[3,4,5,6]

Each of these settings has its own unique characteristics but none as noted by Joorabchi et al[5] are without their shortcomings. New medical students and faculty quickly form their own conceptions and misconceptions of the clerkship programs in the different affiliated hospitals. However, as Schwartz et al[7] found in their evaluation of surgical clerkship experiences, these impressions are more often based upon grapevine influences than objective evaluation data. Few of those involved with clinical education in an affiliated hospital system have accurate data describing specifically how students in different hospital settings acutally spend their time and what impact particular types of experiences have upon the outcomes of the training program. The purpose of this study was to determine if differences in educational experiences did occur in an affiliated hospital system clerkship program and the effect these differences had upon student performance.

*Reprints: Judith G. Calhoun, The Office of Educational Resources and Research, The University of Michigan Medical School, G1111 Towsley, Ann Arbor, Michigan, 48109.

Review of the Literature

A number of studies have been conducted regarding clerkship experiences in affiliated hospitals and site-related measures of students performance. Levitsky,[3] Tamir,[4] Joorabchi,[5] Friedman,[6] and McCarthy[8] all found no differences in measurements of student achievement attributable to affiliated hospital sites. Schwartz et al,[7] however, found significant differences for the entire group of ten affiliated hospitals on three of the four subtests on a final examination designed to test the cognitive objectives of surgical clerkship. Other than Schwartz[7] and Friedman,[6] none of these studies addressed the specific differences intrinsic to the programs and which of these factors may have influenced student scores on the achievement measures.

Schwartz[7] found that one of the hospital sites whose students scored well on the examination was less demanding on students time and, as a result, more time was available for reading and study both at the hospital and at home. In contrast, students assigned to another hospital considered to have one of the better clinical teaching programs by both students and faculty scored much lower. This site was much more demanding of student time for they were occupied with clinical duties must of the day and usually throughout the night when on call every third night.

Friedman et al[6] found that higher percentages of student time were devoted to rounds, interaction with faculty and staff on the hospital floor, and laboratory work; but less time was devoted to clinic activity at the teaching hospital than generally was the case at the community hospitals involved in their study. No significant differences related to the site of the clinical experience were found, however, as measured by a number of standarized test measures.

A time allocation study of student activities by Fisher and Cotsonas[9] addressing the relationship between medical student achievement, as measured by G.P.A., and the allocation of student time across all four years of a medical school curriculum revealed that achievement was not significantly correlated with time spent in formally scheduled activities, study at home, or other study.

Several questions regarding pedagogical and experential differences in clinical education at affiliated hospitals are suggested by the research literature. Similar questions have been posed by students and faculty involved with clerkship programs at The University of Michgian Medical School. Hence, this study was designed to specifically measure and describe the clinical experiences at different hospitals and to determine to what extent intrinsic clerkship differences influence student performance. The specific research questions for the study were: 1) Do students at different hospital sites receive differing clinical experiences?, 2) Do differences in the type of clerkship experiences have an impact on student performance?, 3) Do those students who spend more time on educational types of activities and independent study perform differently than students whose clerkship experiences allow for more practical patient care/clinical exposure?

Method

The study was conducted in the four affiliated hospitals providing a third year (M-3) experience in the Department of Internal Medicine at The University of Michigan Medical School. These four hospitals include: University Hospital, the Veterans Administration Hospital, a county hospital, and a private hospital.

Each of the students (N=233) in the medicine clerkship at the four sites were informed of the study's goals and asked to complete a daily log of their activities for one week near the end of the clerkship. One week in each of the four clerkship rotations was selected for data collection. The instrument for recording all student activities was adapted from one used in a multi-site comparison of clinical education conducted at the University of North Carolina.[6]

The logs contained twenty activity sheets which listed twelve activities and included space to identify four activity variables: 1) the type of activity, 2) the duration of the activity, 3) the disease or condition under discussion or treatment, and 4) the role of the student in the activity. The students were instructed to record their daily activities from the time they entered the hospital until they left the hospital that evening, or the next morning if they were on call.

The twelve clerkship activities were grouped into three categories of activities for analysis (see Figure 1): 1) patient care (activities 1-6), 2) joint patient care and education (activities 7 & 8), and 3) education (activities 9-12). Student role was operationalized as: 1) observed, 2) assisted, 3) performed with assistance, and 4) performed unassisted. The percentage of student time spent in the three activity groups was calculated and analysis of variance techniques were subsequently used to determine activity differences attributable to clerkship site. The data relating to role differences were analyzed by using the Kruskal-Wallis H statistic for rank data. Each role was treated as a rank and the percent of students selecting each rank was determined. These data were then treated as integers for analysis. Product-moment correlations were computed to test for the degree of relationship between the time spent in the three types of activities and student performance as measured by scores on the Internal Medicine Section of the NBME Part II examination.

Figure 1

LOG BOOK ACTIVITIES

PATIENT CARE

1. Initial History & Physical
2. Write-up
3. Chart Work (notes, reviews, etc.)
4. Interaction with hospitalized in-patient
5. Procedure (IV, Lp, etc.)
6. Lab Work

JOINT ACTIVITIES

7. Work Rounds
8. Teaching Rounds

EDUCATION

9. Teaching conference
10. Discussion of medical topic (NOT with
11. Seminar/Lecture
12. Individual Study or Research

Results

Sixty-nine percent (N=800) of the 1165 possible logs were returned. Each log was considered to be a separate case. The analysis of the data for the first research question regarding site-related differences in clinical experiences revealed that both the students' activities and roles in the Internal Medicine clerkship varied significantly (p<.05) among the four sites (Tables 1 & 2).

Table 1

Mean Proportion of Time Per Activity Type by Hospital

Hospital	N	Patient Care		Joint Activities		Education	
	800	%	S.D.	%	S.D.	%	S.D.
Private	170	42.7	17.1	25.3	16.1	31.9	17.3
University	374	45.7	16.4	27.5	15.2	26.8	17.1
VA	138	51.5	15.4	23.7	14.3	24.9	14.5
County	118	50.3	18.2	25.3	14.6	24.4	15.0
Grand		46.7	16.9	26.1	15.2	27.2	16.6
		F=9.27 p<.01		F=2.52 p<.05		F=6.77 p<.01	

Table 2

Proportion of Students Selecting Each Role Level Per Activity Type by Hospital

Hospital	Patient Care Roles				Joint Activities Roles				Education Roles			
	1	2	3	4	1	2	3	4	1	2	3	4
Private	27	42	18	13	2	3	20	75	33	17	15	35
University	29	52	14	5	2	2	9	87	45	13	11	31
VA	28	52	13	7	2	1	8	89	45	14	9	32
County	32	49	10	9	1	1	8	90	47	12	8	33
	H=560		p<.01		H=1251		p<.01		H=510		p<.01	

Students who took their clerkship at the private and university hospitals spent less time on general patient care activities than did the students at the VA or county sites. In addition, the students at the private hospital spent significantly more time on educational activities than the students at the other three sites. When pairwise comparisons were made for each of the 12 activities at the four sites no differences among the sites were found for the following activities: 1) initial history and physical, 2) write-up, and 3) seminar/lecture. Significant site differences were found, however, for the remaining nine activities. As a result, the following characterizations regarding the student activities at the four hospitals were made:

1. The students in the private hospital setting spent a higher percentage of their time interacting with patients and being involved with individual study and research activities, but a lower amount of time on procedures and work rounds as compared to the other three hospitals. These students also spent less time on chart work than did the VA and county hospital students.

2. The students at both the private and university hospital sites reported a higher proportion of time for teaching rounds and a lower proportion for lab work as compared to the VA and county hospitals. These students also reported more time for discussing medical topics than did the students at the VA hospital.

3. The major differences found for the county hospital students as compared to the other hospitals were that they spent a greater amount of time on work rounds with less time in teaching rounds and conferences. In addition, these students spent more time on procedures than did the university hospital students.

In relation to the second research question addressing the impact of differences in clerkship experiences on student performance, the results of the correlational analysis revealed no significant relationships between student performance on the NBME subtest and the proportion of student time spent on the three types of clerkship activities for either the total sample or for each of the four sites (Table 3).

Table 3

CORRELATION BETWEEN INTERNAL MEDICINE NBME PART II PERFORMANCE AND TIME SPENT IN THE THREE TYPES OF CLINICAL ACTIVITIES

Type of Activity	Correlation				
	All Sites	Private Hospital	University Hospital	VA Hospital	County Hospital
Patient	.05	.20	.02	-.08	.04
Joint Patient Care & Education	.04	-.17	.09	.16	.12
Education	-.09	-.02	-.10	-.06	-.15

To determine the relationship between student performance and specific site-related clinical experiences as addressed in the third research question, the correlations for these two dimensions were computed for each of the 12 activities comprising the three types of clinical activities. When all of the sites were considered, a significant correlation was found only for chart work. The only site-related correlations that reached significance at the .05 level were for chart work at the university site and for the discussion of medical topics activity at the county hospital. Neither one of these relationships, however, were for activities in which significant differences in student time were found.

Discussion

Although significant differences in the proportion of time allocated to different clinical activities and the level of responsibility were found in the Internal Medicine Clerkship students at the four hospital sites, the results of the study indicate that these differences are not correlated with performance as measured by the Internal Medicine subtest of the NBME II. Contrary to Schwartz's[7] findings, those students at the clerkship sites allowing more time for educational activities such as teaching rounds, conferences, seminars, independent study and research did not perform any better than the students at other sites which provided a different kind of clerkship experience and distribution of clinical activities. Nor did the students at the hospitals which emphasized patient care activities over educational activities perform differently. Clearly, from these results, it appears to be fallacious for faculty and students to assume that particular hospitals are better sites than others for clinical education based upon the types of clerkship experiences emphasized and the amount of student ime allocated to specific clinical activities. Clerkship experiences may differ markedly and still not affect the outcomes of the training program.

REFERENCES

1. Tally, R. et al. A Modified, Personalized System of Education for Junior Internal Medicine Clerkship at a Multi-Campused School, Proceedings of the 16th Annual Conference on Research in Medical Education, 16:69-76, 1977.
2. Bryne, N. & Cohen, K. Observational Study of Clinical Clerkship Activities, J. Med. Ed., 48:919-927, 1973.
3. Levitsky, S. et al. Are There Educational Advantages to a Surgical Clerkship in a University Hospital?, Surgery, 75:908-914, 1974.
4. Tamir, P., Leshem, H., & Schiffmann, A. Comparative Study of A Pediatric Clerkship in a University & Non-University Hospitals, J. Med. Ed., 49:93-197, 1974.
5. Joorabchi, B., Olson, C., & Page, G. Relationship Between Type of Pediatric Clerkship & Performance of Pediatric Examinations, J. Med. Ed., 48:356-365, 1973.
6. Friedman, C., Stritter, F., & Talbert, L. A Systematic Comparison of Teaching Hospital & Remote-Site Clinical Education, J. Med. Ed., 53:565-573, 1978.
7. Schwartz, G. et al. Evaluation of the Surgical Clerkship Experience in Affiliated Hospitals, J. Surgical Res., 20:179-182, 1976.
8. McCarthy, W. & King, T. Evaluating Patterns of Surgical Clerkship Programs, J. Med. Ed., 43:367-372, 1968.
9. Fisher, L. & Cotsonas, N. A Time Study of Student Activities, J. Med. Ed., 40:125-131, 1965.

PRECIS

PLANNING CONTINUING MEDICAL EDUCATION

MODERATOR: Oscar A. Thorup, M.D.
University of Virginia, School of Medicine

THE EDUCATIONAL VALUE OF A MODEL MEDICAL CARE EVALUATION PROGRAM

A model Medical Care Evaluation program designed specifically to improve the educational value of Medical Care Evaluation activities for physicians who participate in them, is being studied to determine its impact on participants' knowledge, practice, and attitudes. Preliminary results support the educational value of the program, and suggest several features that contribute significantly to the program's educational impact.

CHARACTERISTICS IDENTIFIED UPON ENTRANCE TO MEDICAL SCHOOL ASSOCIATED WITH FUTURE PARTICIPATION IN PROFESSIONAL EDUCATION

This study was an investigation of continuing professional education and those qualities which may guide a professional to be a continuing learner. There exists a significant correlation among the measured characteristics of students entering a professional school and their tendency to becoming continuing lifelong learners.

PHYSICIANS PRACTICE PROFILES: A COMPARISON OF SAMPLING METHODS

A number of popular self-assessment C.M.E. programs base their educational testing and intervention on a practice profile generated from a small sampling of the patient contacts made by each participating physician. The analysis reported here addresses the following question: Do these sampling programs adequately represent the physician's practice profile?

USING MEDICAL AUDIT RESULTS TO PLAN CONTINUING MEDICAL EDUCATION IN COMMUNITY HOSPITALS

The medical audit committee chairmen in each of the 70 general hospitals in central and southern Illinois was interviewed to investigate the usefulness of medical audit results in planning formal continuing medical education activities in their hospitals. Results indicate that even though it appears theoretically sound to base formal continuing medical education programs on needs identified through medical audit, those most directly involved in audit at the local hospital level do not perceive this to be a workable model.

THE EDUCATIONAL VALUE OF A MODEL MEDICAL CARE EVALUATION PROGRAM*

L.J. Sandlow, M.D., P.G. Bashook, Ed.D., J. Maxwell, M.A.
Educational Development Unit
Michael Reese Hospital and Medical Center
Chicago, Illinois

Medical Care Evaluation (MCE) activities are generally regarded as having educational value in two ways: As a means of identifying physicians' educational needs and as a continuing medical education (CME) activity for physicians. Last year, we presented a paper at these meetings criticizing the widely-held view of MCE as educational needs-assessment, and describing the potential CME value of MCE programs like that of Michael Reese Hospital and Medical Center.[1] At that time, we had begun a two-year study to document the impact of a model MCE program that was specifically designed to improve the educational value of MCE for the physicians who participate in it. In the present paper, we present our experiences in implementing this model program, provide some of the preliminary results of the research and discuss the implications of these findings for the design of MCE programs.

RESEARCH DESIGN AND METHODS

A model MCE program, designed along the lines described in our previous publications[1,2], was introduced in three clinical departments (Medicine, Obstetrics and Gynecology and Pediatrics), beginning in March, 1979. The most important features of this program, from an educational point of view, were as follows:

1. The overall purpose of the criteria sets developed through the program is to select appropriate records for review, rather than to set standards.
2. The program focuses on both processes and outcomes of care.
3. Draft criteria are prepared by one member or an invited expert prior to the first meeting on that topic.
4. The focus of criteria committee meetings is on defining appropriate patient care for the problem in question, rather than on the mechanics of criteria formulation: The latter task is handled by the MCE leader and the program coordinator
5. Audit committee members review records prior to the meeting and present the cases for discussion
6. The focus of audit committee meetings is on the overall care of the patient and the issues that this raises, rather than on the criteria not met or the acceptability of care in a narrow sense
7. The intent of the audit committee's actions is investigative rather than punitive

* The research reported in this paper is being supported by a grant from the National Fund for Medical Education. Send reprint requests to Leslie J. Sandlow, M.D., Educational Development Unit, Michael Reese Hospital and Medical Center, 530 E. 31st St, Chicago, IL 60616

8. The committee chair or MCE leader conducts meetings in such a way as to encourage and emphasize educational interchange

9. The committee has a regular attendance of at least four physician members

In each department, the two existing MCE committees (Criteria Development and Audit) were reoriented to the model MCE program and to the purposes of the research. Each committee was to develop criteria or review records for four or five topics (mostly discharge diagnoses) during the research period, which was planned to last for slightly over a year. The committees were to be run by one of the two principal investigators for the project, in cooperation with the committee chair, a physician belonging to the department in which the committee was located. Selection of committee members for the study was by the department chairman for three of the committees, while the other three committees were composed of volunteers.

The research was designed to measure three aspects of the model MCE program: Its effect on participants' knowledge of the topics covered, its impact on participants' patient care practices and the participants' perception of and attitudes toward the program. Knowledge gains were to be determined by using a multiple-choice test on the etiology, diagnosis and treatment of the problems included in the study. Practice changes were to be measured by means of a record audit, using criteria developed for the topics covered by the committees during the research. Perception and attitudes of committee members were to be investigated by using questionnaires, open-ended interviewing and observation and recording of committee meetings.

Several control conditions provide a quasi-experimental design for the research. First, all research techniques (tests, record audit, questionnaires, interviews and observation) were to be used both before and after the research intervention, to provide an accurate measure of changes resulting from the model program. Second, a control group was established for each department and controls were to complete the tests and questionnaires and participate in the record audit. Third, since none of the committees will complete all five topics selected, the remaining topics will constitute a control for the topics covered by the committees. Finally, the test questions and audit criteria were to be separated into two groups: Those dealing with points covered in committee meetings and those pertaining to points not discussed.

RESULTS

The results of the post-tests, record audits and post-questionnaires are not yet available. The following findings and interpretations are based on the initial questionnaires and interview, observation of the committee meetings and anecdotal data obtained from committee members and MCE staff and on the authors' experiences in implementing the model program:

The implementation of the model MCE program was not equally successful in all departments (see Table 1). We found that if the MCE leader is not a physician, as in the committees in Departments B and C, there can be serious difficulties in establishing and maintaining the focus required by the model. In these two departments, the MCE leader only partially achieved the goals of the model program. In Department A, on the other hand, where a physician served as both MCE

Table 1: Comparison of Committees with Model Program

Criteria Development Committees	A	B	C
Preparation of draft criteria	+	+	+
Emphasis on patient care	++	+	0
Educational emphasis	++	+	0
Regular attendance of ≥ 4 physicians	++	+	0

Audit Committees	A	B	C
Presentation of previously-reviewed records	++	+	+
Emphasis on overall care and problem	++	0	0
Educational emphasis	++	0	0
Regular attendance of ≥ 4 physicians	++	0	+

++: Model completely or always met
+: Model partially or sometimes met
0: Model minimally or rarely met

leader and chairman, there was far less difficulty in operating the committees according to the model.

We also found that there was a tradeoff between MCE value and educational value of the committee meetings. Too much emphasis on MCE tasks resulted in a low educational value, as described in our previous paper[1]. However, too much emphasis on educational activity decreases the committee's MCE efficiency and can lead to dissatisfaction and loss of interest by committee members, though this depends to some extent on how successfully the educational aspects of the model program are implemented.

The results of the questionnaire administered at the beginning of the committees' participation in the research provide some useful data on the perceived educational value of the different committees, and thus on the effect of conformity to, or deviation from, the model MCE program. Those committees that resembled the model most closely (Criteria Development Committee A and Audit Committee A) were considered the most educational, while the committee that was considered the least educational (Criteria Development Committee C) was one that was relatively far from the model (see Table 2). Particularly significant was the fact that when respondents were asked how educational the committee had been for them personally, those who had served on Audit Committee A gave that committee a higher rating than they had when asked how educational it was for its members in general, while for the other committees, the ratings from the two questions were approximately the same.

The frequency of identified learning opportunities occurring during committee meetings during the research period provides another test of the educational value of the model program (see Table 3).

Far more significant than these quantitative differences, however, are the qualitative differences in the educational interchange in the committees. In Criteria Development Committees A and B, and in Audit Committee A, there is

Table 2: Members' Perceptions of Committees' Educational Value

Criteria Development Committee	A	B	C
Very	2	2	1
Moderate	2	8	1
Slight	0	2	1
None	0	0	2
	4	12	5

Audit Committee	A	B	C
Very	3	4	0
Moderate	5	5	1
Slight	1	4	0
None	0	0	0
	9	13	1

Table 3: Number of Learning Opportunities per Meeting

Criteria Development Committee	A	B	C
Mean	11.3	5.6	3.0
S.D.	6.4	4.6	2.0

Audit Committee	A	B	C
Mean	6.9	3.5	5.0
S.D.	2.1	3.5	1.8

often a high-level discussion of the issues surrounding a problem or patient, with citation of research findings. In Audit Committee C, on the other hand, the learning that occurs is often of a low level, with Audit Committee B and Criteria Development Committee C being intermediate in educational value.

There are also several types of observational data confirming the educational nature of the committees in Department A, which most closely approach the model program. These committees are normally characterized by a high level of attention, interest and participation during the meetings. Members frequently ask questions about the appropriate management of the problem or about the case being reviewed; in many cases, it is clear from the content of the question or from direct statements by the member that the question is being asked for their own information, rather than being relevant to criteria development or record review. This interest is reflected in the actions of committee members in Department A with respect to committee meetings. In Criteria Development A, members decided to continue meeting to develop criteria for another problem, although the four topics needed for the study had been completed and the committee was under no obligation to continue. In Audit Committee A, in which members are appointed to one-year terms by the Department Chairman, two members have asked to continue on the committee after their period of service ended, citing its interest and educational value for them. A third member, an older physician with an important position in the hospital administration, specifically requested appointment to this committee "as a necessary refresher," and, when interviewed, spoke at length about how educational the committee had been for him. Another older physician in an administrative position regularly attends meetings of Criteria Development Committee A, and participates actively in the discussion, although he is not formally a member of the committee.

There is also a certain amount of direct evidence of learning. In several meetings of Criteria Development Committee A, members have taken notes during the discussion. Some members have also specifically mentioned things that they learned from the committee's deliberations. We expect that this direct evidence will be greatly increased when the interviews, knowledge tests and audit results following the research period are analyzed.

DISCUSSION

There are several key features of the model MCE program that appear to be primarily responsible for its greater educational value. For Criteria Development committees, two points are most important. First, at least one person should be involved who is an authority on the topic, either a committee member or an invited expert. Without this, discussions frequently get bogged down in fruitless argument with no effective resolution.

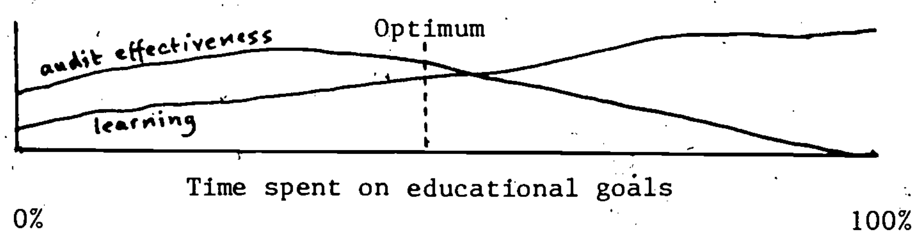
Second, the focus of the committee should be on the problem and its diagnosis and treatment, and not on standards, deviations and exceptions. This is not an easy task for the MCE leader or chairman to accomplish, particularly in a committee that is used to the second approach. The most effective technique is to ask the committee, not what standards or criteria should be set, but how a patient with this problem should be managed, and inferring the criteria from the resulting discussion. The criteria-setting process thus constitutes a step-by-step review of the management of such a patient, from the findings needed to support the diagnosis to the expected outcomes of treatment. The MCE leader focuses the committee's attention on the problem by concretizing the issues with hypothetical examples, using such phrases as "Now you've admitted the patient with a diagnosis of _____; what do you do next?" In this way, the Criteria Development Committee becomes a conference on the health problem in question, exploring its nature, diagnosis, treatment and outcomes. One meeting of Criteria Development Committee C was conducted by a physician in the manner described, and a far greater number of learning opportunities occurred than was usual for this committee.

For Audit Committees, two other characteristics of the model MCE program are most important. First, records are reviewed prior to the meeting and the case is presented by the reviewer for committee discussion. When records are reviewed during the meeting, or are not presented as cases, educational discussion is minimal. Second, the focus of the review, presentations and discussion should be on the overall care of the patient and the issues that this raises, and not narrowly restricted to the criteria that were not met or the acceptability of the care. Again, this is not easy to implement, and has been completely accomplished only in Audit Committee A. However, the educational difference between the two approaches is striking to an observer.

In terms of committee structure, several variables are significant. First, the size of the committee has an effect on its educational value; committees with fewer than four physician members tend to be less educational than those with four or more members. Second, it seems to be important to the effective implementation of the model MCE program that the meetings be led by someone who is both a physician and has some experience in education. A non-M.D. will have difficulty guiding the discussion and gaining committee members' acceptance of his role; a non-educator, unless thoroughly familiar with the program goals and techniques for attaining them, with have difficulty keeping the discussion educational unless the committee is capable of maintaining this focus wihtout overt leadership.

IMPLICATIONS

We have presented evidence in this paper that an MCE program can be a significant educational experience for the physicians that participate in it, and have identified some of the factors that contribute to the educational value of such a program. However, we have also found that there is a tradeoff between education and quality assurance in such a program. Too much emphasis on education detracts from the tasks of criteria development and record review, and vice versa. However, our experience has been that up to a certain point, the goals of education and evaluation are not only compatible, but reinforce one another. This can be diagrammed as follows (figure 1):



Some attention to educational goals increases physician interest and commitment to the committee's work, and results in deeper and more relevant investigation of health problems and patient care, producing better criteria sets and more adequate record review. Studies at Michael Reese Hospital and Medical Center have shown that criteria sets developed by the process described above are effective in identifying those records exhibiting questionable care.

Finally, the principles developed here for making MCE programs more educational should be generalizable to other types of programs dealing with health problems or patient care. These principles include the value of focusing on management of the patient and allowing standards or committee actions to develop out of this; the importance of having the leader or facilitator be a physician; and the non-linear relationship between educational goals and committee tasks.

REFERENCES

1. Sandlow, L.J., M.D. et al. "Medical Care Evaluation: An Experience in Continuing Medical Education." Proceedings of the Eighteenth Annual Conference on Research in Medical Education. Association of American Medical College, November, 1979, Washington D.C.

2. Sandlow, L.J. et al. Patient Care Evaluation: The Implementation of a Model System. Michael Reese Hospital and Medical Center, Chicago, Illinois, 1979.

CHARACTERISTICS IDENTIFIED UPON ENTRANCE TO MEDICAL SCHOOL ASSOCIATED WITH FUTURE PARTICIPATION IN PROFESSIONAL EDUCATION

Linda K. Gunzburger, Ph.D., Director
Division of Continuing Medical Education
Loyola University Stritch School of Medicine

This study was an investigation of continuing professional education and those qualities which may guide a professional to be a continuing learner. Both continuing general adult learning activities and continuing professional learning activities were studied (1).

Purpose

The purposes of this study were to respond to the following areas of questions (2):

1) Is it possible to develop a technique to measure the extent a professional participates in continuing professional education activities?

2) Is it possible to identify characteristics of incoming professional school students which may indicate the extent to which the students will later participate in continuing learning activities?

3) Is there a positive correlation between the amount and quality of time devoted to continuing learning carried out in leisure time activities and the amount and quality of time devoted to continuing professional learning activities which may occur during leisure or work time?

Other studies have not considered if any of the facts collected about entering medical students are valid predictors of the extent each student will later be a continuing professional learner. In studying students at the University of California, School of Medicine, San Francisco, Doctor Gough examined the Medical College Admission Test science subtest score, premedical grades in science courses, and a preference index for science courses. Gough concluded that students having high scientific aptitude do superior work early in medical school, but by year four are indistinguishable from their classmates with regard to clinical competence. By rating students on knowledge,

---

Send Requests for Reprints to:
Linda K. Gunzburger, Ph.D., Director
Division of Continuing Medical Education
Loyola University Medical Center
2160 S. First Ave., Maywood, IL 60153

data-gathering skills, clinical judgment and professional attitudes, it was found at the Jefferson Medical College that future performance in the first postgraduate year may be predictable. No significant relationship was found between MCAT Science subtest and ratings in clinical competence (3).

Methodology

This study concerned professionals who had entered professional school over twenty years ago. By dealing with such a group of individuals it was possible to determine the extent each participates in continuing professional education activities and the extent each participates in continuing adult learning activities. The following procedures were used for this study:

a) The incoming professional school variables for each individual were defined and related to the measure of his current continuing professional learning.

b) The amount of time a professional devotes to continuing general adult learning participation was compared to the amount of time devoted to continuing professional learning. If professionals tended to spend a similar proportion of time participating in both professional education and in continuing adult learning activities, then the time devoted to continuing professional learning and continuing adult learning could be considered as related.

c) The quality of continuing professional learning and the quality of the continuing general adult learning were compared.

Sample

This study examines the 1956 University of Chicago School of Medicine incoming class. The 1956 freshman class was part of a national longitudinal study sponsored by the Association of American Medical Colleges (AAMC) (4).

Design

Data for the study were collected during two separate time periods, 1956 and 1978. In 1956 entering students completed the Medical College Admission Test, the Allport-Vernon-Lindzey Study of Values, the Edwards Personal Preference Schedule, and demographic questionnaires. It was the 1946-1962 form of the Medical College Admission Test that was administered to the University of Chicago sample and used in this study. This Medical College Admission Test had four subtests. These subtests were Verbal Ability, Quantitative Ability, General Information, and Science (5). The Allport-Vernon-Lindzey Study of Values considered theoretical, economic, aesthetic, social, political, and religious values (6). The Edwards Personal Preference Schedule measured achievements, deference, order, exhibition, autonomy, affiliation, intraception, succorance, dominance, abasement, nurturance, change, endurance, heterosexuality, and aggression. The demographic questionnaires requested demographic information frequently asked on medical school applications (7).

The Leisure Activity Survey, Activities Survey, and interviews were used to collect data from the sample in 1978. The Leisure Activity Survey (LAS) measures the nature and number of general adult learning activities during leisure time while the Activities Survey (AS) measures the continuing professional learning activities of physicians during work time or leisure time (8). The Activities Survey was constructed to measure the extent physicians participate in continuing medical education during leisure or work time. The Activities Survey provides each physician a Continuing Medical Education Index which identifies the extent of his involvement in Continuing Medical Education activities. An interview questionnaire was developed to provide each individual the opportunity to discuss specific childhood experiences and the effect such experiences may have on his later continuing professional education activities; continuing medical education programs currently available; present continuing medical education activities; and the use of vacation time. The interviews were conducted after the LAS and AS were completed.

Conclusions Regarding the Hypotheses

Hypothesis I states: <u>There exists a significant correlation among the measured characteristics of students entering a professional school and their tendency to becoming continuing lifelong learners.</u> For purposes of this study, a significance level of 5 percent (sig. = .05) is accepted. Because of the importance of certain other results, findings not significant are also discussed. First, it was found the four MCAT subscores all showed a negative relationship with the Activities Survey score. Although the strength of the relationship did not reach the required level of significance, two subscores, Verbal Ability ($r = -.29$, sig. = .09) and Modern Society ($r = -.30$, sig. = .08) approached significance. Furthermore, it was found that of the three instruments given to entering medical students two subscores on the EPPS showed significant correlations with Activities Survey scores representing continuing professional education. The subscores, sig. $\leq .05$, were Change ($r = .49$) and Heterosexuality ($r = .31$). Two other EPPS subscores, Autonomy ($r = -.26$, sig. = .10) and Aggression ($r = -.26$, sig. = .10) fell below the required level of significance but showed strong negative relationships with the dependent variable, continuing professional education. When the incoming test subscores of the MCAT, Allport-Vernon-Lindzey Study of Values, and EPPS were correlated with the Leisure Activity scores, no significant relationships were found.

Second, demographic information frequently found on completed medical school applications was studied to determine if any correlations existed between the demographic variables and later evidences of continuing learning. It was found that the level of father's education and the level of mother's education correlated .29 and .32 with the Activities Survey, significant to the .04 and .025 levels, indicating increased levels of parental education correlates with higher Activities Survey scores; high school class rank correlates -0.35 with the Leisure Activity Survey, significant to the 0.017 level, indicating that higher Leisure Activity Survey scores are obtained by persons with lower high school class rank.

Third, after the Activities Survey and Leisure Activity Surveys were completed by each physician, a sampling of thirty-six physicians was interviewed. Those individuals who could remember using the library weekly as a youngster had higher Activities Survey and Leisure Activity Survey scores. As adults, flexible individuals who enjoyed new and adventuresome activities tended to have higher Activities Survey and Leisure Activity Survey scores. Individuals who did new and different things were also identified as continuing learners. These individuals also had a higher Edwards Personal Preference Schedule Change subscore.

Fourth, multiple regression analyses were used to analyze the 1956 and 1978 data and to derive two predictor equations. The 1956 incoming test subscores are defined as independent variables and the Activities Survey and Leisure Activity Survey scores are defined as dependent variables. Consequently, the following equation, significant to the .005 level ($r^2 = 0.40$, $r = 0.63$), may be used to predict the Activities Survey score. Activities Survey score = 116. (EPPS Change) + 67. (EPPS Heterosexuality) - 104. (EPPS Aggression) - 103. (EPPS Nurturance) + 11396. For any incoming student it is possible to substitute his EPPS Change, EPPS Aggression, EPPS Heterosexuality, and EPPS Nurturance subtest scores into the predictor equation. The solution to the equation is the projected Activities Survey score.

In a similar manner the best LAS predictor equation uses the Edwards Personal Preference Schedule Change subscore as the only predictor element and is: LAS = 2. (EPPS Change) + 112. r^2 is equal to 0.08, $r = 0.28$, and is significant to the 0.10 level.

Based on the significant correlations of the incoming subtest scores with the Activities Survey and Leisure Activity Survey scores, the demographic data, interview findings, and the continuing learning predictor equations, Hypothesis I was accepted.

The second hypothesis states: There is a positive correlation between the amount of time devoted to continuing learning carried out in leisure time activities and the amount of time devoted to continuing professional learning as carried out in professional time activities. To test the second hypothesis the average values of the time weights for the AS were compared to the average values of the time weights for the LAS for each physician. These mean values of the time weights for the Activities Survey and Leisure Activity Survey are correlated at 0.53 with a significance level of 0.00005. This means that it is highly probable that an individual involved in continuing adult learning activities during leisure time will also devote a similar proportion of time to continuing professional learning during leisure and work time. Also, a person involved in continuing professional learning activities during leisure and work time will devote a similar proportion of time to continuing adult learning during leisure time.

The third hypothesis states: There is a positive correlation between the quality of continuing learning carried out in leisure time activities and the quality of continuing professional learning as carried out in professional time activities. The quality (intensiveness) of learning is a synthesis of both the time spent at an activity and the judged educativeness of the activity. To test the third hypothesis it was necessary to compare the Activities Survey and the Leisure Activity Survey scores. The correlation between the Activities Survey and the Leisure Activity Survey scores is .41 (sig. = .002). This significance level and correlation suggest there is a positive relationship between the Activities Survey and Leisure Activity Survey scores, and Hypothesis III was accepted.

Implications of the Study

From the Chicago sample, it is now possible to identify characteristics which describe those qualities which may guide one to be a continuing professional learner. Such information should be most useful to admissions officers at professional schools. Since it is desirable for a professional to be a continuing learner and since no measurable evidence exists that professional school requirements inquire if an applicant will be a continuing professional learner, it is worthwhile to expand the scope of this study to other medical schools. The expanded study can include other medical schools that participated in the 1956 Association of American Medical Colleges longitudinal study (8).

Limitations

This study is limited by dealing with one profession and a single professional school. Although much of the entrance data requested was similar at all professional schools, different schools as well as different professions do have varying entrance criteria. For this reason, the results of this study, although generally applicable to all professions at all schools, were specifically oriented to one sample group of one profession at one particular professional school.

A Concluding Note

One value of this study was developing the system for studying the problem. Perhaps this systems approach can be applied in an expanded project. The various requirements of the different professional schools indicate a means to identify an intellectually able student; however, few, if any, questions asked of an applicant can reveal life long learning characteristics. The study did reveal characteristics which an admissions committee may choose to review in questioning if an applicant will become a life long learner.

References

1. Houle, C.O. The Inquiring Mind. Madison, Wis.: University of Wisconsin Press, 1961.
2. Houle, C.O. The Lengthened Line. Perspect. Biol. Med., 11:40, 42, 1967.
3. Veloski, J., Herman, M.W., Gonnella, J.S., Zeleznik, C., and Kellow, W.F. Relationships Between Performance In Medical School and First Postgraduate Year. J. Med. Educ., 54:909-916, 1979.
4. Hutchins, E.B. Profiles on Selected Student Variables from the AAMC Longitudinal Study of the Class of 1960. Evanston, Illinois: Division of Education, Association of American Medical Colleges, 1969.
5. Erdmann, J.B., Mattson, D.E., Hutton, J.G., Jr., and Wallace, W.L. The Medical College Admission Test: Past, Present, Future. J. Med. Educ., 46:939-9[illegible], 1971.
6. Allport, G.W., Vernon, P.E., and Lindzey, G. Manual for Study of Values. Cambridge: Riverside Press, 1951.
7. Edwards Personal Preference Schedule Manual. New York, New York: The Psychological Corporation, 1953.
8. Litchfield, A. The Nature and Pattern of Participation in Adult Education Activities. Unpublished Ph.D. dissertation. University of Chicago: Department of Education, 1965.
9. Ad Hoc Committee on Continuing Medical Education of the Association of American Medical Colleges. Continuing Education of Physicians: Conclusions and Recommendations. J. Med. Educ., 55:155, 1980.

PHYSICIANS PRACTICE PROFILES: A COMPARISON OF SAMPLING METHODS*

Lynn Curry, Ph.D.**
Dalhousie University, Faculty of Medicine

A number of popular self-assessment C.M.E. programs base their educational testing and intervention on a practice profile generated from a small sampling of the patient contacts made by each participating physician. The accuracy of this profile limits the accuracy of the learning needs identification for the physician and therefore the efficiency of the educational program. The analysis reported here addresses the following question: Do these sampling programs adequately represent the physician's practice profile?

BACKGROUND

The interest in self-assessment and individualized C.M.E. has been growing, and has resulted in a number of intervention programs that utilize individual physician practice profiles as a basis to identify educational needs. The relationship between the profile, the physician self-assessment and the remediation offered depends on the particular intervention program.

Among the first was the "Individualized Physician Profile" package, offered by the University of Wisconsin, Department of Continuing Medical Education, developed in 1968. This program has three phases, the first of which is the practice profile. Participating physicians are asked to complete information "on each patient contact for a different day each week for four weeks". Based on this information, a profile of the participant's practice is developed and represented in ICDA (International Classification of Diseases, Adapted) codes. A self-test of 125 multiple choice questions is generated at the same time, with more questions from heavily represented practice areas. The participant completes the test and discusses test performance, the profile, attitudes and office management with a CME faculty consultant. This process specifies educational need. The intervention concludes when the C.M.E. faculty consultant mails the participant a computer listing of educational opportunities and events in those areas of priority need. The I.P.P. program has been described in some depth by Sivertson et al (1973, 1974).

A similar program is regarded sufficiently well by the American Medical Association that the organization has purchased distribution rights. The College of Physicians of Philadelphia PREP (Practice Related Educational Program) program instructs participants to complete an information form "for each of the first 100 patient contacts that you have during your participating days". Participants are cautioned to choose "those days of the week which are most typical of your practice". This collection of information is optically scanned and a practice profile based on

---

* This paper is based on research funded by Health and Welfare Canada, NHRDP grant #603-1075-44

** Requests for reprints should be sent to Lynn Curry, Ph.D., Assistant Director for Research, Division of Continuing Medical Education, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7.

simplification of ICDA returned to the participant by mail. Presumably based ~~on the profile information, the participant then chooses one of the thirteen~~ specified topic areas for a pre-instruction self-assessment test of various multiple choice questions. The test results are returned to the physician along with the list of available learning materials in the specific areas chosen by him. The requested items are drawn from the College's library and mailed to the physician. After sufficient time has elapsed, the physician completes another test to determine the success of the learning experience. Bowler et al (1977) have described the PREP program more fully.

Following these models, the College of Family Physicians of Canada (CFPC) is currently developing a profile based self-assessment package. One hundred pilot participants have been asked to record information on 200 consecutive patient encounters. This information will be optically scanned and a profile generated using a modified version of ICHPPC codes (International Classification of Health Problems in Primary Care). The profile will be returned to the participants along with a multiple choice examination of 200 questions. Of these, 150 questions will be directly based on profile areas of high incidence. In this program educational intervention is left to the participant upon receipt of examination results.

All of these programs attempt to insure individual relevance, a fundamental educational principle, by making use of the participant's practice profile. This utilization is not trivial. In all cases, the self-assessment examination and hence the educational intervention is predicated completely (IPP) or to a large measure (PREP, CFPC) upon the observed and reported profile. It should be obvious that to produce true relevance and applicability, the accuracy of the observed profiles should be examined.

Each of these programs uses a small sample of patient encounters in order to estimate a physician's practice profile. There is a large question about the accuracy of these small samples. How well do the various kinds of sampling approximate the true practice profile?

METHOD

Our research group was in a unique position to investigate this question as part of an extensive research project into the effects of physician participation in the process of patient care appraisal. Sixteen participating physicians were asked to keep continuous log sheets of patient contacts in the office setting for two six-month periods. A review of available literature on physician profiles indicated that this baseline is one of the most extensive currently available, in terms of both number of physicians and length of time (Curry and MacIntyre, 1980).

Using the profiles generated from the first six months of contact coding, we conducted a series of analyses to duplicate the sampling in each of the described individualized C.M.E. programs: 100 consecutive contacts, 200 consecutive contacts, all contacts for four different days in four different weeks. Only six months of data were used in this comparison in order to reduce the complexity of the analysis and because this amount yields a sufficiently large number of patient encounters to allow the necessary comparisons. The average number of encounters per physician was 3,516; ranging from 2,640 to 4,419.

For the analysis of the 100 contact method used by the PREP group, we took the first 100 contacts each physician recorded each week for the duration of their collection period which ranged from 25-32 weeks. This yielded an average of 28 separate samples of 100 consecutive patient contacts for each physician. A similar procedure was followed for the 200 contact method used by CFPC, yielding an average of 27 separate samples for each physician. For the "one day per week for four weeks" sample, the IPP method, we first eliminated the one day per (5-day) week that each physician took as time off. Where the day off was not clearly indicated, we arbitrarily chose Wednesday to eliminate. The next step was to collect for each physician all patient contacts for Monday the first week, Tuesday the second week, Thursday the third week and Friday the fourth week. This was repeated throughout the six month collection an average of 7 times for each physician, with sample sizes ranging from 273 to 710 patient contacts.

ANALYSIS

We began with a chi-square comparison of the ICHPPC frequencies observed in each sampling to the frequency that would be expected if the whole six month pattern was taken as the standard. This comparison was done separately for each physician and for each of the three sampling methods. The resulting patterns of chi-squares were examined using Pearson's test of uniformity (Rao, 1965) to determine if the physician's practice, as indicated by these small samples, varied more than would be expected by random fluctuation. This might occur for example, if the sample was too small and the practice had a cyclic fluctuation inside each week.

RESULTS

1. For the 100 consecutive contact method.

The results displayed in Table 1 indicate that this method of sampling was adequate for only two of the 16 physicians. In one further case, the adequacy of the sampling was rejected at the 2.2% level. In all other cases the evidence against the representativeness of the sampling was overwhelming.

2. For the 200 consecutive contact method.

It was hoped that taking larger sized samples would weaken the erratic fluctuations about the expected frequencies that were observed with the samples of size 100. Unfortunately, the fluctuations were merely reinforced when the sample size doubled. It is clear from Table 2 that selecting 200 consecutive cases does not, in general, give an adequate view of the profile category frequencies over a longer run. For Doctor #10, the sampling was adequate, and for three others the sampling adequacy is rejected at the 5.6% level. For the remaining physicians, however, the sampling adequacy decreased markedly from the already poor showing in the 100 consecutive sample method.

3. For the one day per week method.

This method is adequate for three physicians, somewhat close for three others, but completely wrong in the remaining ten physicians (Table 3).

DISCUSSION AND CONCLUSIONS.

This analysis indicates clearly that the current methods of sampling a physician's practice are statistically inaccurate when compared to a six month continuous profile. The practical significance of this inadequacy hinges on where the importance is placed in the C.M.E. intervention process predicated upon these profiles. Sampling inaccuracy will lead to some profile areas being under-represented on the examinations, and some over-represented. This in itself is no great sin, but the next step is indefensible, that poor performance in a high profile area is critical to correct. The reliability of that profile peak is cast into serious doubt by the here demonstrated inadequacy of all three current sampling methods. If, however, all that is important in these self-assessment programs is that the physician feel that his examination and resulting C.M.E. has been tailored to his practice profile, then the accuracy of that profile is not critical.

We suspect that the program developers would prefer to have an accurate profile, but are constrained by the time and effort they can ask of practising physician participants. For that reason, we are proceeding to examine our twelve month data base to find the minimum accurate sample size.

149

PHYSICIAN #	$(-2)\sum_j \ln P_j$	D.F. Two times the number of samples	P-value
1	118.223	62	0.0026×10^{-2}
2	74.640	52	0.0215
3	84.357	54	0.0052
4	92.954	56	0.0014
5	102.761	56	0.0001
6	54.729	52	0.3714**
7	110.141	58	0.0005×10^{-1}
8	99.503	52	0.0001
9	82.780	52	0.0042
10	57.194	50	0.2256**
11	95.481	54	0.0004
12	102.763	54	0.0001
13	96.695	64	0.0011
14	140.481	56	0.0004×10^{-5}
15	80.510	50	0.0040
16	126.019	62	0.0032×10^{-3}

** Samples not significantly different from standard, ($\alpha \geq .06$) and therefore sampling is adequate for that doctor.

TABLE 1. Samples of 100 consecutive contacts compared to the complete record (six months' continuous coding).

PHYSICIAN #	$(-2)\sum_j \ln P_j$	D.F. Two times the number of samples	P-value
1	140.213	60	0.0003×10^{-4}
2	69.190	52	0.0559
3	71.514	54	0.0559
4	123.175	56	0.0007×10^{-3}
5	139.851	54	0.0002×10^{-5}
6	66.878	50	0.0556
7	135.509	58	0.0005×10^{-4}
8	109.350	50	0.0003×10^{-2}
9	112.866	50	0.0001×10^{-2}
10	63.347	50	0.0973**
11	103.770	52	0.0003×10^{-1}
12	108.083	52	0.0009×10^{-2}
13	107.852	62	0.0029×10^{-1}
14	271.484	54	0.0003×10^{-25}
15	107.983	48	0.0002×10^{-2}
16	131.074	60	0.0004×10^{-2}

** Samples not significantly different from standard ($\alpha \geq .06$) and therefore sampling for that doctor is adequate.

TABLE 2. Samples of 200 consecutive contacts compared to the complete record (six months continuous coding).

PHYSICIAN #	(-2) $\sum_j \ln P_j$	D.F. Two times the number of samples	P-value
1	42.786	16	0.0003
2	19.459	14	0.1481**
3	25.255	14	0.0321
4	51.127	16	0.0015 X 10^{-2}
5	38.553	14	0.0004
6	24.228	14	0.0430
7	40.968	16	0.0006
8	34.025	14	0.0020
9	28.070	14	0.0139
10	19.712	14	0.1392**
11	41.589	14	0.0001
12	51.362	14	0.0036 X 10^{-3}
13	26.484	16	0.0476
14	128.712	14	0.0001 X 10^{-6}
15	32.501	14	0.0012
16	21.874	16	0.1473**

** Samples <u>not</u> significantly different from standard ($\alpha \geq .06$) and therefore sampling is adequate for that doctor.

<u>TABLE 3.</u> Samples of one day per week for four weeks compared to the complete record (six months continuous coding).

REFERENCES

1. Bowler, F.L.; Brading, P.L.; Burg, D.F.; Finestone, A.J.; Hubbard, J.P. A Practice-Related Educational Program. J.A.M.A. 237(13):1346-1349, 1977.

2. Curry, L.; MacIntyre, K. A Comparison of Profiling Techniques. (In prepartion for publication).

3. Fisher, R.A. Statistical Methods for Research Workers. 13th Edition. Oliver and Boyd, 1963. pp. 99-101.

4. Rainsberry, P. College of Family Physicians of Canada, Personal communications. November and December 1979.

5. Rao, C.R. Linear Statistical Inference. John Wiley and Sons, New York, 1965. Section 3a.3.

6. Sivertson, S.E.; Meyer, T.C.; Hansen R.H.; Schoenenberger, A.O. Individual Physician Profile: Continuing Education Related to Medical Practice. Journal of Medical Education 48:1006 - 1012, 1973.

7. Sivertson, S.E.; Hansen, R.H.; Shropshire, R.W.; Schoenenberger, A.O. Family Practices in Wisconsin: Implications for Medical Education and Delivery of Health Care. Wisconsin Medical Journal 73:S170-S174, 1974.

Using Medical Audit Results To Plan Continuing Medical Education in Community Hospitals*

Charles E. Osborne, Ed.D., Assistant Dean
for Continuing Medical Education
and Associate Professor of Medical Education
Southern Illinois University School of Medicine

Introduction and Purpose

With the limited time available to practicing physicians which can be spent in formal CME, the selection of pertinent topics becomes critical. The first step in organizing any effective CME activity is needs identification since CME is generally believed to be more effective when directed toward specific problems in patient care which have been pinpointed by the medical audit (1-6). Both the JCAH and the PSRO program endorse this approach (7-10).

A discussion of the use of audit results and the necessity for CME based on medical audit is presented by Brown (11) and Sanazaro (2). Recent estimates of the percent of problems identified by the medical audit which are related to lack of physician's medical knowledge and possibly correctable by a formal CME program range from less than one to fifteen percent (12, 6). There are also data to indicate that physician knowledge is not necessarily related to actual performance (2, 11, 13, 14). Even though the literature contains many reports of improvement in medical care attributable to CME (11, 15-17), there is a lack of scientific documentation of the existence of benefits resulting from formal CME programs (18). There are also data questioning the relationship between audit results and actual physician performance (2, 14, 19-23).

Mason and Kappelman (24) reported that 62% of the CME programs that medical school CME directors considered to be "successful," were designed to meet physician needs as identified by the organizers of the sessions, while only 28% addressed needs identified by the participants. In only two instances were programs designed to meet needs identified by chart audit of hospital or office records. Since medical school sponsored CME programs are designed for physicians from many varied settings, the results of any one hospital's audits are of little meaning. There needs to be a reason to believe the problem is generalizable to other settings before these data are meaningful in planning CME for an audience from many different hospitals. However, when a CME program is designed for the staff of one hospital, audit results of that hospital's performance should provide useful planning data for education programs. This study was designed to evaluate the perception of hospital medical audit committee chairmen regarding the role and usefulness of medical audit results in planning their hospitals' CME activities.

Methods

An interview protocol was developed and field tested. Questions were open-ended to keep from leading chairmen's responses. To help assure the consistency of the investigation, one researcher in the Office of Continuing Medical Education conducted all interviews. The audit committee chairman in each of the 70 general hospitals in central and southern Illinois was interviewed. Each chairman was

* Requests for reprints should be addressed to Charles E. Osborne, Ed.D., Office of Continuing Medical Education, Southern Illinois University School of Medicine, P.O. Box 3926, Springfield, Illinois 62708

asked questions related to 1) the purpose of their hospital's audits, 2) how topics are determined, and 3) their committee's competence at a) developing, b) conducting, and c) following-up on the results of their audits. Questions also investigated the usefulness of audit results in planning formal CME activities and what other sources of information were used to plan its CME activities.

Sixty-five of the hospitals had hospital audit committees serving all departments within the hospital. Two of the 65 hospitals shared a joint medical audit committee. The other five hospitals had departmental audit committees. In those hospitals with departmental audit committees, each committee chairman was contacted. If any chairman refused to be interviewed, he was offered the opportunity to complete a written questionnaire containing the same information. For purposes of analysis, the hospitals were divided into three classifications according to number of beds in each hospital. There are 33 (47%) small hospitals which contain less than 100 beds; 21 (30%) medium hospitals with 100 to 200 beds; and 16 (23%) large hospitals with 201 or more beds. One medium sized hospital and four large hospitals have department audit committees. In one of those four large hospitals, each departmental audit chairman reported to the Medical Care Committee, whose chairman was interviewed for this survey.

Results

Audit chairmen from 67 of the 70 hospitals agreed to the telephone interview. One chairmen completed a written questionnaire. The responses from the committee which was responsible for audit in two hospitals were only counted once in the final analysis, but each hospital was counted as having responded. Results from the 13 department audit committee chairmen (representing the three responding large hospitals and one medium hospital with department committees) and those of the 62 hospital audit committee chairmen interviewed were not significantly different in any of the areas investigated (x^2, $p < .05$). Therefore, for most reporting purposes, their answers were combined and results reported on 75 interviews: 32 chairmen from small hospitals; 18 from medium hospitals; 12 from large hospitals; and 13 department chairmen.

Seventy-nine percent of the chairmen responded that the JCAH requirements were at least partially responsible for their hospital conducting audits. The two most common other reasons given for performing audits were: 1) to improve the quality of patient care and 2) to identify problem areas. The JCAH was perceived as the only reason their hospital conducts audits by twice as many audit committee chairmen in small hospitals as in either medium or large hospitals (60% compared to 30%). If audits were not required by the JCAH, 75% of chairmen from medium and large hospitals indicated that they would continue to conduct audits, compared to less than 50% from small hospitals.

Committees in approximately one-half of the small and large hospitals with department audit committees generally selected audit topics based on number of patients admitted in a given diagnosis group. Eighty percent of all hospitals used either the most common diagnosis or left the decision to members of the audit committee. In very few instances were audits conducted in diagnosis areas where the committee had received information that there was a problem with the care being delivered. However, there is no way to know whether individual committee members used this reason when suggesting topics. Several chairmen commented that they believed this to be the case but could not be sure. The primary method(s) used to select audit topics are presented by hospital classification in Table 1.

Only 19% of those surveyed disagreed with the statement, "audits are a good ~~method of identifying needs upon which formal CME programs can be based" (Table 2).~~ However, only 23% could identify a CME program which had been developed from audit data. Formal CME programs based on audit results were more likely to be presented in large hospitals (42%) or by departments (31%) than in either medium (11%) or small (19%) hospitals. When a deviation from the pre-established standards was identified, a formal CME program generally did not result. The most common occurrence was that audit results were discussed at staff/department meetings and recommendations of changes in medical procedures (or their documentation) were adopted (23%) or a letter was sent to the doctor(s) whose performance differed substantially from the pre-set criteria (16%). Other actions listed included: changing the criteria used in the audit, recommending changing hospital procedures, encouraging more elaborate chart documentation, purchasing new equipment, and doing nothing.

When the audit did not result in a formal CME program, eighty percent of the time one of the following three reasons was cited: 1) the chairman did not have the time or inclination to put on formal CME programs for an identified discrepancy which was traceable to one or two physicians with whom the chairman could either discuss the problem or initiate a formal letter from the Chief of Staff, 2) the problem identified was administrative or logistical rather than related to a lack of physician's medical knowledge, and could be managed by a change in hospital procedures and, 3) too little information was available from the audit upon which to base a CME program. The other 20% of the responses were spread among the following reasons: too busy to develop a formal CME program, no facilities to provide formal CME, did not know how to set up a formal CME program, not authorized to present formal CME programs, and "I don't know."

The chairmen's rating of their committee's performance on each of the three stages of the medical audit (the development of criteria, the actual conducting of the audit, and the follow-up on the results of the audit) is presented in Table 3. There was not a significant difference in the perception of the chairmen regarding their committees performance on the three stages (x^2, $p < .05$). Most chairmen believed the success of the committee was directly related to the performance of the medical records personnel.

Seventy-eight percent of the chairmen believed that CME should be based on the interests of individual physicians and on new techniques rather than audits. Three other responses which accounted for the remaining 22% were: update of new material, audits, and problems identified in patient care.

Discussion

Even though it appears theoretically sound to base formal CME programs on needs identified through medical audit, those most directly involved in audit at the hospital level in central and southern Illinois do not perceive this to be a workable model. Particularly in smaller hospitals, the problems identified by medical audit are not currently remedied by formal CME. Nevertheless, since only 3% believe they are doing a poor job in following-up on audit results, the performance of audits may be resulting in informal education activities which are, in every sense, educational for the staff and may improve the quality of patient care. The fact that audit results do not always lead to formal CME programs may also be due in part to the belief that audits do not accurately measure quality of care, and, therefore, do not provide an accurate assessment of needed programs.

The majority of small hospitals complete medical audits because they are required to do so by the JCAH. Small hospitals find audit to be a less productive tool than large hospitals. There is the belief that the JCAH requirements have created unnecessary extra paperwork for hospital personnel, especially those in the Medical Records Department. In smaller hospitals it is believed to be relatively easy to identify the one or two physicians responsible for any identified deviation from expected standards. To develop formal CME programs for the whole staff to reach these one or two persons is not viewed as either cost effective or an efficient use of physician time. As one chairman stated, "In our hospital the staff knows each others strengths and weaknesses; the audit is an expensive way to confirm them." Many medium and large hospitals would continue to conduct audits even if the requirement were lifted by the JCAH. Perhaps in the larger hospitals with more support staff the burden of conducting medical audits is spread among more people reducing the time needed from any one person. In the small hospital, staffing the audit committee is more likely to disrupt the Medical Records Department since the responsibility commonly falls on one person.

A program of self or individualized instruction may be a method of dealing with the one or two staff members for whom a particular need is identified. Building educational programs to meet an individuals need, while possibly not cost effective for the provider of the CME experience, may in the long run prove to be quite cost effective for the health care delivery systems since those who do not need the program are practicing medicine rather than receiving instruction in areas where they have already demonstrated competence.

Medical audits, even though perceived by audit committee chairmen to be well designed and well conducted, are not generally detailed enough to provide information upon which to base a formal CME program. Although they may identify areas where CME is needed, the content of the audit seldom provides enough detailed information upon which to develop the content of a formal CME program. The 1980 Edition of JCAH's Accreditation Manual for Hospitals deals with this dilemma by recognizing alternative mechanisms for assessing the quality of patient care.

Many audits identify no deviation from expected standards and thus there is no need for formal or informal CME. One reason for this is that good quality care is being delivered. Two possible explanations for the proportion of audits that reveal no discrepancies could be that quality care is generally being practiced or standards have been set too low. Another possible explanation could be that the areas being audited are not those where problems are likely to be identified. Hospitals which choose audit topics by frequency of diagnosis, at random, or because criteria have previously been developed by another agency, may be failing to capitalize on the possible benefits of audit. Auditing areas where outcomes have been poorer (e.g., high patient mortaility, morbidity, complication, or readmission rates) than expected, or where the audit committees' judgement tells them there may be problems, could result in a larger percentage of audits documenting existing problem areas.

The author recommends that future research develop and investigate more cost effective means of identifying needs upon which formal CME programs can be based. The fact that audit results do not lead to formal CME programs is reasonable. Many other more informal educational activities may be shown to well serve the overall purpose of quality assurance programs...the continuing education of physicians and the improvement of patient care.

References

1. Brown, C.R. Jr., Fleisher, D.S. The bi-cycle concept--relating continuing education directly to patient care. In: Stearns, N.S., Getchell, M.E., Gold, R.A., eds. Continuing education in community hospitals. Boston: Postgraduate Medical Institute, 88, 1971.
2. Sanazaro, P.F. Medical audit, continuing medical education and quality assurance. West J Med, 1976; 125;241, September.
3. Stein, L.S. Runkel, N. CME planning checklist. Illinois Council on Continuing Medical Education, Chicago, February 1978.
4. Williamson, J.W., Alexander, M., Miller, G.E. Continuing education and patient care research: physician response to screening test results. JAMA 1967; 210:938.
5. Williamson, J.W., Alexander, M., Miller, G.E. Priorities in patient care research and continuing medical education. JAMA 1968; 204:303.
6. Workshop of CME leadership (syllabus). Illinois Council on Continuing Medical Education, Chicago, May, 1976.
7. Joint Commission on Accreditation of Hospitals. Accreditation manual for hospitals. 1979 ed. Chicago.
8. Goran, M.J., Roberts, J.S., Kellogg, M. et al. The PSRO hospital review system. Med Care 1975; 13(4) (supplement).
9. Jessee, W.F., Munier, W.B., Fielding, J.R. et al. PSRO: an educational force for improving quality of care. N Engl J Med 1975; 292:668.
10. Joint Commission on Accreditation of Hospitals. The PEP primer. Professional evaluation procedure for auditing and improving patient care. Chicago 1974.
11. Brown, C.R. Jr., Uhl, H.S.M. Mandatory continuing education: sense or nonsense? JAMA 1970; 213:1660.
12. Jessee, W.F. Quality assurance systems--why aren't there any? Paper presented at the Association for Health Records Eighth Annual Multidisciplinary Conference on Health Records. Washington, D.C. June, 1977.
13. Gonnella, J.S., Goran, M.J., Williamson, J.W. et al. Evaluation of patient care: an approach. JAMA 1970; 214:2040.
14. Osborne, C.E., Thompson, H.C. Criteria for evaluation of ambulatory child health care by chart audit; development and testing of a methodology. Pediatrics (supplement) 1975; 56:625.
15. Devitt, J.E. Does continuing medical education by peer review really work? Can Med Assoc J 1973; 108:1279.
16. Payne, B.C. Continued evolution of a system of medical care appraisal. JAMA 1967; 201:536.
17. Williamson, J.W., Aronovitch, S., Simonson, L. et al. Health accounting: an outcome-based system of quality assurance: illustrative application to hypertension. Bull NY Acad Med 1975; 51:727.
18. Bertram, D.A., Brooks-Bertram, P.A. The evaluation of continuing medical education: A literature review. Health Educ Monogr 1977; 5:330.
19. Osborne, C.E. Interdiagnosis relationships of physician recording in ambulatory child health care. Med Care 1977; 15:465.
20. Storey, P.B. Mandatory continuing medical education. One step forward-two steps back. N Engl J Med 1978; 298:1416.
21. Stross, J.K., Harlan, W.R. The impact of mandatory continuing medical education. JAMA 1978; 239:2663
22. Thompson, H.C., Osborne, C.E. Office records in the evaluation of quality of care, Med Care 1976; 14:294.
23. Thompson, H.C., Osborne, C.E. Quality assurance of ambulatory child health care: opinions of practicing physicians about proposed criteria. Med Care 1976; 14:22.
24. Mason, J.L., Kappelmen, M.M. A survey of medical school continuing education programs. J Med Educ 1977; 52:341.

TABLE 1. Number and Percent of Audit Committees in Each Hospital Classification Using Various Methods to Select Audit Topics

Primary Method of Topic Selection*	Hospital Classification				
	Small	Medium	Large	Departmental Committees	Total
Most Common Diagnosis	17 (53%)	6 (33%)	4 (33%)	7 (54%)	34 (45%)
Suspected Problem Area	8 (25%)	3 (17%)	2 (17%)	1 (8%)	14 (19%)
At Random	3 (9%)		1 (8%)	1 (8%)	5 (7%)
Committee Discussion (e.g., each member suggests a topic)	9 (28%)	12 (67%)	7 (58%)	4 (31%)	32 (43%)
Criteria Already Available	1 (3%)	2 (11%)	1 (8%)	3 (23%)	7 (9%)
None Conducted Prior to 7/78	1 (3%)				1 (1%)
Number of Interviews	32	18	12	13	

* Several of those interviewed responded with more than one primary reason; therefore, % will add to more than 100.

TABLE 2. Number and Percent of Audit Committee Chairmen Who Agreed With the Statement "Audits are a Good Method of Identifying Needs Upon Which Formal CME Programs Can Be Based"

Hospital Classification		Response		
		Agree	Neutral	Disagree
Small	n = 32	17 (53%)	9 (28%)	6 (19%)
Medium	n = 18	10 (56%)	3 (17%)	5 (28%)
Large	n = 12	9 (75%)	1 (8%)	2 (17%)
Departments	n = 13	11 (85%)	1 (8%)	1 (8%)
TOTAL	n = 75	47 (63%)	14 (19%)	14 (19%)

TABLE 3. Percent of the 75 Chairmen Interviewed Rating the Performance of Their Committee Between Excellent and Poor

Stages of Audit	Rating				
	Excellent 1	2	3	4	Poor 5
Development of Criteria	14%	43%	41%	1%	1%
Conducting the Audit	15%	51%	30%	4%	
Follow-up on Results	14%	39%	33%	11%	3%

PRECIS

IMPACT OF RESIDENCY PROGRAMS: MEASUREMENT PERSPECTIVES

MODERATOR: Kenneth W. Rowe, Jr., M.D.
University of Cincinnnati, School of Medicine

FACULTY PERCEPTIONS OF AMERICAN AND FOREIGN PEDIATRIC RESIDENTS

Differences between faculty ratings of foreign and American graduates are described along three dimensions -- character, cognition, and consultation seeking, and the educational implications of these differences are discussed.

PERFORMANCE ON PART III OF THE NATIONAL BOARDS - THE EFFECT OF RESIDENCY TRAINING

It was determined that scores on Part III of the National Board Examinations are influenced by the type of postgraduate training program. The graduates who enrolled in internal medicine, family medicine and flexible programs performed at a higher level than those in five other programs.

PATIENT INSTRUCTORS AS EVALUATORS OF HOUSESTAFF CLINICAL COMPETENCE

Symptomatic non-physician patient instructors (PIs) with stable describable findings on physical examination have been trained to function independently of physicians as patients, teachers, and evaluators of houseofficers' physical diagnosis skills. They are able to provide objective data which correlate significantly with intership acceptance committee rankings, program director ratings, and peer evaluations.

FACULTY PERCEPTIONS OF AMERICAN AND FOREIGN PEDIATRIC RESIDENTS*

T. Joseph Sheehan, Ph.D.,** Susan D. R. Husted, B.S., Department of Research in Health Education, University of Connecticut Health Center, Farmington, CT; Dan Candee, Ph.D., Center for Moral Education, Harvard University, Cambridge, MA; Charles D. Cook, M.D., Department of Pediatrics, Downstate Medical Center, Brooklyn, NY.

Introduction

The presence of foreign-born residents and physicians practicing in the United States is considerable. Until recently one-half of all physicians licensed each year to practice medicine in the United States were born and trained abroad. They constitute 20% of all physicians in the United States. Thus, it is of great importance in terms of planning and evaluating residency programs to understand the ways in which "foreign" residents are perceived by faculty supervisors as being similar to or different from their American counterparts.

Background

Although there is a general impression that foreign medical graduates perform more poorly than their U.S. counterparts (Nile, 1976, Williams, 1975), there is little evidence that these perceptions translate into measurable differences at the attending level (Saywell, 1979). Hospital effect rather than physician characteristic seems to explain more of the variance in performance by the time physicians reach the attending level. Perceived differences at the house-staff level are important, since there is a long series of screening mechanisms before the resident becomes an attending: state licensure examinations, analysis of credentials before receiving hospital privileges, formal and informal peer review (Saywell, 1979). At the house staff level the foreign graduates are likely to be experiencing language difficulty and cultural shock, which may influence the way they are judged by faculty supervisors.

Weis (1974) has shown differences between U.S. and foreign medical graduates on medical knowledge as measured by the E.C.F.M.G. examination. Studies by Margulies, Bloch and Cholko (1968) and by Halberstam and Dacso (1965) reach somewhat contrary conclusions. While both studies report differences, the Margulies study warns of a potential disastrous effect of such differences on standards of practice, while Halberstam warns that many perceived differences may simply be the result of one self-fulfilling prophecy feeding on another rather than any real differences in performance.

The current study is intended to describe perceived differences in performance and to discuss the implication of such differences for house staff education.

Methods

The 157 residents in this study were chosen from five different university-related training programs. Sixty Americans were selected from three programs, and 97 foreign graduates from the remaining two. Neither the selection of programs nor residents was random. However, the programs did represent a wide range of quality as viewed by professional colleagues (Sheehan, et al, 1980).

*The research reported here was based at the University of Connecticut Health Center, Farmington, and is supported in part by grants from the National Fund for Medical Education, the Commonwealth Fund, and the University of Connecticut Research Foundation.

** Requests for reprints should be directed to Dr. Sheehan.

From two to five faculty supervisors rated each resident using a rating form describing twenty aspects of clinical performance plus a rating of overall performance. The rating form was adapted from items validated by Cook and Margolis (1974) and cast into a semantic differential format for each of the twenty performance characteristics. Each item was rated on a seven-point scale ranging from 1 (positive pole) to 7 (negative pole). The rating of overall performance was based upon a four-point scale. Cook and Margolis (1974) report reliabilities in the .75 range. For a subsample of 26 residents rated by four common faculty members, the reliability of the mean rating ~~for overall~~ performance was 0.86, while the average intercorrelation among four raters was 0.67 (Sheehan, et al, 1980).

Further studies on the construct validity of the performance rating scale (Candee, et al, 1980) reveal three underlying latent variables or factors. The first factor is characterized primarily by interpersonal and character items: relates well to patients, works well with others, is compassionate, admits mistakes, is honest, and knows own limits. The second factor seems to deal primarily with cognitive characteristics: organization, fund of medical knowledge, decisiveness, clinical judgment, and seeks medical knowledge. The third factor is characterized by empathy and seeking consultations. These three factors, when taken together, account for 75 percent of the variance shared by all twenty of the items on the rating scale.

Results

Table 1 presents the mean ratings for foreign and American residents. Foreign graduates have poorer ratings on all twenty performance characteristics, including overall performance. Due to the direction of the scale these poorer ratings are indicated by higher mean scores. All of the differences are statistically significant, and all but two differences are at or beyond the .001 level of significance. The largest differences appear on the qualities of "empathy," "seeks clinical consultation," "is personable," "is compassionate," "is honest," "relates well to patients," and "seeks medical knowledge."

Table 2 summarizes differences between foreign and American graduates which are computed from weighted combinations of performance items combined into the three factors found by Candee, i.e., character, cognitive and consulting. The mean scores for each factor are presented at the bottom of Table 2, along with differences between the means for each of the two groups. While the absolute difference in means is relatively uniform across the three factors the magnitude of the difference relative to the means is considerably larger for factor 3.

Discussion

First, it should be stressed that these results are based upon faculty ratings which are undoubtedly influenced by factors other than actual performance. The ratings are subjective judgments based upon faculty perceptions.

It may not be surprising that foreign graduates are rated lower than American graduates on clinical performance. The foreign graduates are rated slightly worse in overall performance, a mean of 2.3 for the foreign graduates versus a mean of 1.9 for the Americans. All twenty items on the subscale reflect this same difference favoring the Americans. It is the profile of the differences that attracts attention.

On the first factor (character and interpersonal) the biggest differences between foreign and American graduates are seen on the items dealing with honesty, compassion, and relating to patients. The foreign graduates are seen as less honest, less compassionate, and poorer in relating to patients.

On the second factor (cognitive) the biggest differences between foreign and American graduates involve medical knowledge, the seeking of medical knowledge, and clinical judgment. Each of these differences is close to a full standard deviation. Foreign graduates are seen as knowing less about medicine, as being less active in seeking medical knowledge, and as having poorer clinical judgment.

The differences on the third factor may be the most interesting and of the most practical value. This factor contains the two items showing the largest absolute difference between the two groups, empathy and seeks medical consultation. The foreign graduates are seen as less empathetic and as seeking medical consultation less than the Americans. The reason for the added importance of these differences is that they may be the most tractable. Differences between the groups on the cognitive factor may simply reflect previous educational experience. Research and previous efforts indicate that changes in knowledge are fairly easy to affect. One would expect that with continued training and sufficient motivation, knowledge differences could eventually be minimized.

In terms of the second set of differences, the prognosis is less clear. Less is known about affecting change in honesty, compassion, or relating to patients. The same would be true of empathy which appears in the third factor. However, there seems to be great potential to affect changes in resident consultation-seeking behavior. In fact, it would even seem that improvement in seeking consultations would partially compensate for perceived weaknesses in medical knowledge and clinical judgment. This would seem to be a reasonable policy to implement in residency education.

Conclusions

Significant and important differences between foreign and American residents, as measured by faculty ratings, can be clustered into three summary categories: character, cognitive and consultation seeking behaviors.

Differences on at least some of the qualities composing two of the factors, cognitive and consultation seeking seem easily amenable to change. Admittedly, our discovery of considerable differences in supervisors' perceptions of the use of consultations by foreign and American residents is not equivalent to having discovered that the groups do, in fact, use consultations differently. However, we suggest that programs which teach foreign residents to make more appropriate use of consultations might go a long way towards improving their communications with American residents and supervisors and also in assuring a more uniform standard of medical care in this country.

TABLE 1

MEAN RATINGS OF FOREIGN AND AMERICAN PEDIATRIC RESIDENTS

	Mean			
	Foreign	American	Δ	t*
Organized	3.14	2.75	.39	2.53
Admits Mistakes	2.99	2.65	.34	2.51
Accepts Responsibility	2.78	2.33	.45	3.26
Honest	2.60	1.98	.62	5.17
Dependable	2.76	2.29	.47	3.07
Works Hard	2.77	2.32	.45	3.29
Personable	3.45	2.65	.80	5.84
Relates Well to Patients	3.18	2.51	.67	4.89
Compassionate	3.34	2.50	.8[illegible]	6.13
Empathy	2.95	1.30	1.65	12.05
Works well with Others	3.00	2.57	.43	3.14
Medical Knowledge	3.35	2.78	.57	3.93
Teaches Well	3.68	3.25	.43	2.44
Seeks Medical Knowledge	3.13	2.40	.73	5.21
Aware of Own Limits	3.12	2.75	.37	2.68
Technical Skills	3.10	2.81	.29	2.04
Decisive	3.25	2.76	.49	2.99
Clinical Judgment	3.19	2.64	.55	3.80
Seeks Consultations	2.72	1.33	1.39	5.81
Responds Well to Emergencies	3.27	2.73	.54	2.83
Overall	2.3	1.9	.40	4.57

*All significant at P <.001, except two.

TABLE 2

MEAN FACTOR SCORES OF FOREIGN AND AMERICAN PEDIATRIC RESIDENTS

	Mean			
	Foreign	American	Δ	t*
Factor 1, Character	13.2	10.9	2.3	4.16
Factor 2, Cognition	12.5	10.4	2.1	3.82
Factor 3, Consultation	5.0	2.3	2.7	6.43

*All t values significant at P<.001.

REFERENCES

Candee, D., Sheehan, T.J., Husted, S.: Stability of factors in the evaluation of residents' performance. (In preparation)

Cook, C.D., Margolis, C.Z.: Rating pediatric house officer performance. Pediatric Research, 8:472, 1974.

Halberstam, J.L. and Dacso, M.M.: Foreign and United States residents in university-affiliated teaching hospitals. An investigation of United States graduate medical education. Bull. NY Acad. Med. 42:182, 1966.

Margulies, H., Bloch, L.S., and Cholko, F.K.: Random survey of U.S. hospitals with approved internships and residencies. A study of the professional qualities of foreign medical graduates. J. Med. Educ. 43:706, 1968.

Nile: Behind the squeeze on the FMGs. Hosp. Phys. 16-19, 37, January 1976.

Saywell, R.M.,Jr., Studnicki, J., Bean, J.A., and Ludkes, R.L.: A performance comparison: USMG-FMG attending physicians. AJPH January 1979, Vol. 69, No. 1.

Sheehan, T. J., Husted, S., Bargen, M., Candee, D., Cook, C. D.: Moral judgment as a predictor of clinical performance. Evaluation and the Health Professions (in press).

Weiss, R.J., Kleinmann, J.C., Brandt, U.C., et al.: The effect of importing physicians-- return to a pre-Flexnerian standard. N.Engl.J.Med. 290:1453, 1974.

Williams, K.A. and Brook, R.H.: Foreign medical graduates and their impact on the quality of medical care in the United States. MMFQ: 549-581, Fall 1975.

5/80
/se

PERFORMANCE ON PART III OF THE NATIONAL BOARDS - THE EFFECT OF RESIDENCY TRAINING

Jon Veloski, M.S., and Joseph S. Gonnella, M.D., Office of Medical Education
Jefferson Medical College

While scores on Parts I and II of the National Board Examinations (NBE) have been used to monitor curricular changes, similar use of Part III has not been emphasized.[1,2] The clinical competence of graduates is an important measure of the quality of medical education.[3] Because Part III of NBE is taken by most graduates soon after completion of medical school, the results of this examination could be used as one measure of the physicians' competence and as a reflection of the effectiveness of the educational programs.

Although the empirical validity of NBE Part III and similar examinations has been questioned [4,5], the content and face validity of this examination lend support for its present use in the absence of an alternative. Consisting of patient management problems and multiple choice questions based on clinical material in written and pictorial form, it is intended to measure the clinical competence of physicians in training as the final step in a process of certification. Since Part III attempts to sample broad areas of clinical knowledge and skills required of all physicians, and since the examinees have spent the nine months prior to the examination in a wide variety of postgraduates programs, it is possible that scores might vary according to the type of program of each graduate. No reports of such differences have been published to date.

In this study we tried to determine whether or not performance on the Part III examination is influences by the type of postgraduate program taken in the first year after graduation from medical school (PGI). We expected that graduates who followed the more general training programs such as internal medicine, family medicine or flexible (rotating in earlier years) would achieve higher scores than those who entered programs which lead to earlier specialization, such as surgery, obstetrics/gynecology or psychiatry. Since the groups might differ in levels of knowledge and skills prior to residency training, scores on Part II of the NBE were used for statistical correction of these differences.

Method

The data for the present study of 1564 graduates between 1970 and 1978 were derived from a longitudinal study of the graduates of the Jefferson Medical College. Data were excluded for 213 graduates who followed a 5 year BS-MD program. The remaining physicians were classified according to the type of postgraduate program followed in the first year. Eighty-seven graduates who pursued certain specialized programs, such as anesthesiology, radiology and urology, were not included in the present study since the number of graduates in any one of the programs was too small for meaningful interpretation. Of the 1264 graduates remaining, scores on Part III were available for 1028 (81%) graduates who had given permission for the medical

Reprints: Jefferson Medical College, 1025 Walnut Street, Philadelphia, Pennsylvania 19107.

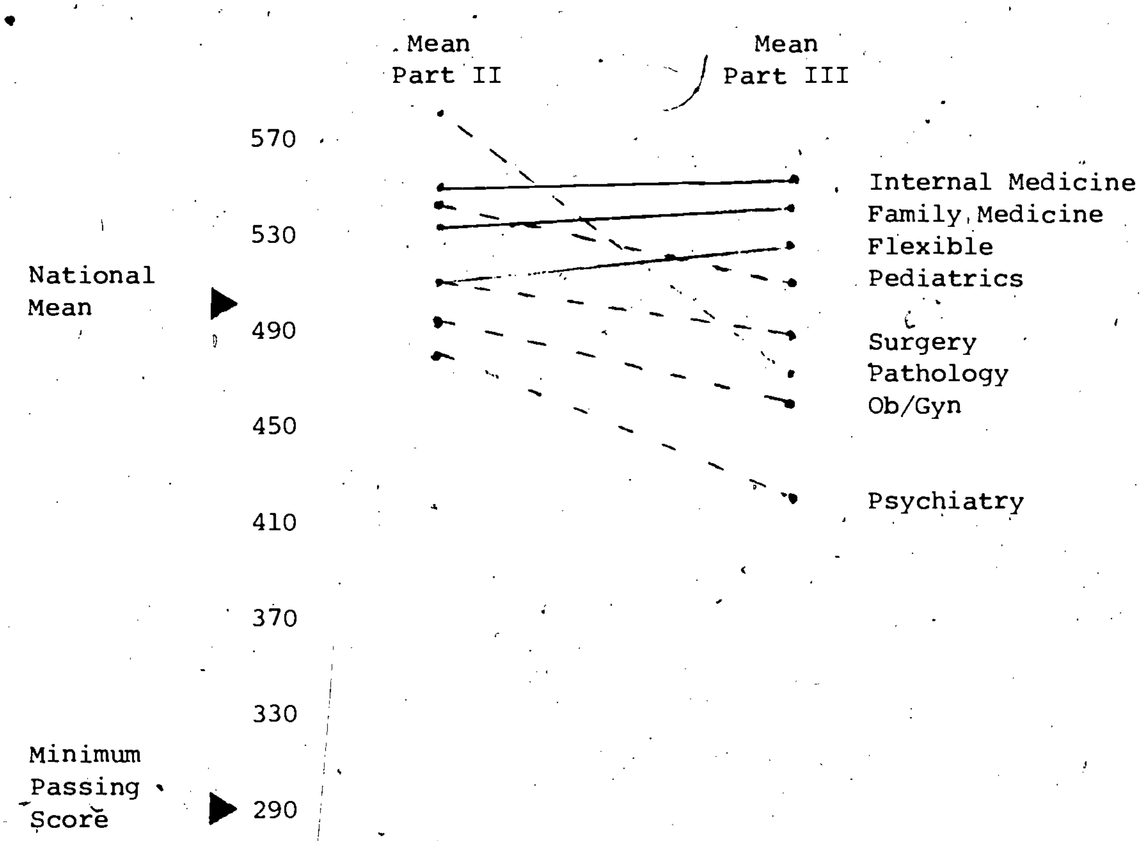
school to obtain reports of their scores on Part III. A chi square test comparing the distribution of the types of PGI programs of those with scores (n = 1028) to those without scores (n = 236) detected no differences ($x^2 = 10.75$, df = 7, $p < .15$) with regard to the type of program. Since scores on NBE Part II were available for all graduates the mean of Part II total scores for those with Part III scores was compared to the mean for those who had not given permission. The mean score for Part II was greater ($p < .01$) for those who gave permission ($\bar{x} = 535$) than for those who did not ($\bar{x} = 515$).

The hypothesized effect of the type of postgraduate program on Part III scores was tested with a one-way analysis of covariance (ANCOVA) where the adjustment for prior differences was based on each subject's total score on NBE Part II. The homogeneity of the within groups regression lines was tested and not rejected ($p < .60$). The Neuman-Keuls procedure was used to compare the means on Part III, after adjustment for Part II.

Results

Diagram 1 shows the changes in mean scores from Part II to Part III for the graduates in eight PGI programs. Table 1 shows the means and standard deviations of scores for these programs, and Table 2 presents the means on Part III adjusted by the results of the ANCOVA. The most noticable declines are seen in pathology and psychiatry, while levels of performance

Diagram 1 CHANGES IN MEAN SCORES FROM PART II TO PART III FOR GRADUATES IN EIGHT POSTGRADUATE PROGRAMS



156

Table 1

MEANS AND STANDARD DEVIATIONS
PARTS II AND III OF THE NATIONAL BOARD EXAMINATIONS
WITHIN TYPE OF POSTGRADUATE PROGRAM

Program	n	Part II	Part III
Family Medicine	111	532 88	540 98
Flexible	221	511 88	512 97
Internal Medicine	328	553 91	554 96
Ob/Gyn	46	495 90	467 96
Pathology	25	574 100	479 99
Pediatrics	75	543 88	501 92
Psychiatry	24	485 77	432 92
Surgery	198	512 83	492 85
All Subjects	1028	531 93	519 70

ANALYSIS OF COVARIANCE - PART III

	df	ss	F	p > F
Part II	1	4064138	835	.0001
PGI Program	7	481426	14	.0001
Error	1019	4959215		

are relatively unchanged for graduates in family medicine, internal medicine and flexible. The table for the ANCOVA of NBE Part III scores shows a statistically significant effect for PGI after adjustment for Part II.

Comparisions of the adjusted means using the Neuman-Keuls procedure at alpha = .10 identified three groups of PGI programs. One group includes family medicine, internal medicine and flexible. The second is comprised of obstetrics/gynecology, pediatrics and surgery. The third is psychiatry and pathology.

Table 2 MEAN SCORES ON PART III OF THE NATIONAL BOARD EXAMINATIONS ADJUSTED FOR PERFORMANCE ON PART II BY ANCOVA WITHIN POSTGRADUATE PROGRAM

Program	n	Adjusted Mean Part III
Family Medicine		538
Internal Medicine		536
Flexible		525
Surgery		504
Pediatrics		492
Ob/Gyn		492
Psychiatry		464
Pathology		447

Note: The brackets indicate programs grouped according to tests for differences by the Neuman-Keuls procedure. Means within each group are not significantly ($p < .10$) different.

Discussions and Conclusions

Part III of NBE is a measure of clinical competence and is used not only for licensure but also for the evaluation of the product of medical education. In this study we determined whether performance on Part III is influenced by the type of PGI program. We expected and found that the graduates who entered the broader, less specialized programs performed at a higher level than those in other programs. The sample was representative of the graduates who entered the eight types of PGI programs, but tended to include graduates with records of higher performance on the NBE Part II examination.

These results have implications for medical colleges, the graduates and the postgraduate training programs. While it may be argued that those who pursue pathology as a career may need knowledge and skills that are different than those measured by Part III, the same cannot be said for the other groups which showed measurable declines. If scores on Part III are, in part, a function of the type of PGI program taken by the physicians, this variable should be recognized if the results are used as a measure of the quality of the undergraduate programs. In addition, variation in scores may also be due to the faculty and unique features of each hospital's envionment, and of course the characteristics of the graduates who select different programs.

Graduates whose scores on Part II of the NBE are marginal should be counseled about the potential influence of their postgraduate experience on their performance on Part III. In the present study the performance of nearly all graduates is far above the minimum passing level for Parts II and III, but this may not be true for other samples of graduates. However, in other samples the academically marginal graduate who is entering a PGI program leading to early specialization might be advised to undertake independent study to prepare for Part III.

One of the most important consequences of these results must be addressed by the postgraduate programs themselves. If Part III does in fact measure a subset of clinical knowledge and skills expected in all physicians, these findings suggest that these competencies are not being fostered equally in all programs. Even if Part III were replaced by the Comprehensive Qualifying Exam Program (CQEP) or the Federation Licensing Examination (Flex I and II) its present format and content will influence these new examinations of clinical competence. Therefore, if our findings are documented by others, our results provide support for a broader general education in the clinical sciences before a physician enters speciality training as recommended by The Council on Medical Education of the AMA, in its report "Future Directions for Medical Education".[6]

REFERENCES

1. Shapiro, A.P., Schuck, R.F., Schultz, S.G., and Barnhill, B.N. The Impact of Curricular Change on Performance On National Board Examinations. J. Med. Educ. 49:1113, 1974.

2. Levit, E.J. Use of the National Board Minitest for Evaluation of Curricular Change. J. Med. Educ. 42:930, 1967.

3. Veloski, J., Herman, M.W., Gonnella, J.S., Zeleznik, C., and Kellow, W.F. Relationships between Performance In Medical School and First Postgraduate Year. J. Med. Educ. 54:909, 1979.

4. Goran, M.J., Williamson, J.W., and Gonnella, J.S. The Validity of Patient Management Problems. J. Med. Educ. 48:171, 1973.

5. Mackowiak, R. and Gonnella, J.S. Clinical Competence As Measured By The National Boards, Part III. Proceedings of the Conference on Research in Medical Education, 1973.

6. "Future Directions for Medical Education". A Report of The Council on Medical Education - AMA, December, 1979.

PATIENT INSTRUCTORS AS EVALUATORS OF HOUSESTAFF CLINICAL COMPETENCE

Paul J. Rutala, M.D., Paula L. Stillman, M.D.,
and Darrell L. Sabers, Ph.D.
(University of Arizona College of Medicine, Tucson, Arizona)

In 1970, the American Board of Internal Medicine (ABIM) discontinued the oral examination as part of its certification process. Formal evaluation of housestaff by the ABIM is now accomplished by written examinations which contain multiple choice questions and patient-management problems. These assess a candidate's factual knowledge and ability to diagnose and manage clinical problems when specific historical, physical, or laboratory data have been provided. While these examinations have proven useful in the assessment of the candidate's cognitive abilities, they bear little relation to actual clinical encounters in which the physician must integrate and synthesize data he himself has collected. In addition, these methods of evaluation provide little information regarding the candidate's attitude, his skill in developing rapport with patients, or his technical ability in eliciting a history and performing a physical examination.

The responsibility for assessing the board-eligible internist with regard to these actual clinical skills has now been delegated to the individual programs which train internal medicine housestaff. Evaluation methods are often subjective and dependent upon the directors of the training programs. Input is often sought from many faculty members, each of whom may use somewhat different criteria to judge candidates.

In an effort to obtain a somewhat more objective evaluation of a candidate's clinical performance, the ABIM has provided training programs with a form which can be completed by a faculty member as he observes a houseofficer examine an actual patient (the "fully observed patient encounter"). Limitations of this method are many: the physician-patient encounter may be affected by the presence of the faculty observer, the method requires a considerable commitment of faculty time, the physical examination findings detected by the houseofficer cannot be verified unless the faculty member re-examines the patient himself, and an element of subjectivity remains which is dependent upon the observer and individual patient chosen for the encounter. This evaluation may suffer from the same lack of reliability which was one of the major reasons the oral examinations were themselves discontinued (1). Further, an on-site ABIM survey of 161 training programs between 1973 and 1975 revealed that only 21% of the programs conducted a "one-to-one" exercise where a trainee was observed during an entire encounter (2).

Videotaped physician-patient encounters have been used for evaluation purposes in some programs. Although the evaluator need not be present in

---

The work reported here was supported in part by grants from the National Fund for Medical Education and the Department of Health, Education, and Welfare.

Reprint requests to: Dr. Rutala in the Section of General Internal Medicine, Arizona Health Sciences Center, 1501 North Campbell Avenue, Tucson, Arizona 85724.

the room at the time of the examination, the videotape must be reviewed later. Despite the fact that feedback to the houseofficer after videotaping leads to increased completeness of the physical examination, the improvement in performance does not ensure improved identification of pathology or correct interpretation of findings (3). Further, certain maneuvers of the physical examination are subjective and an evaluator reviewing a videotape may not accurately discern if they have been done. Only the examining physician and patient directly involved in the encounter can judge such maneuvers.

Other training programs have advocated the use of chart audit to measure the performance of houseofficers in clinical situations (4). Using computers and non-physician abstractors may conserve physician time and increase the amount of data available on each trainee. However, the accuracy of physical findings described in the chart are not always verified by a second observer and the houseofficer's attitudes and relationships with patients are not assessed.

A new objective method for assessing competence in basic clinical and diagnostic skills has been developed at the University of Arizona College of Medicine (UACM) (5). This new strategy overcomes many of the deficiencies inherent in other evaluation models by utilizing trained non-physicians as evaluators. This program has been adapted to evaluate housestaff.

THE PROGRAM

Seven chronically-ill patient instructors (PIs) with stable, abnormal cardiovascular or pulmonary findings were trained by physicians to evaluate the physical diagnosis skills of first-year housestaff (PGY-1s). The PIs were judged by representatives of the internal medicine faculty to have physical examination findings that all houseofficers should recognize. Their diagnoses included asthma, bronchitis, bronchiectasis, mitral regurgitation, click-murmur syndrome, and aortic stenosis with aortic insufficiency. The PIs were taught to evaluate the thoroughness and proficiency of the houseofficer's examination and to ascertain whether the houseofficer correctly identified and described the abnormal findings.

All PIs had been employed previously as part of a program to evaluate the physical diagnosis skills of second and third year medical students at UACM (5). They were hired as staff members of the University and paid an hourly wage. After the PIs had conducted more than 600 sessions with three classes of medical students, the program was adapted to evaluate three groups of PGY-1s: twelve internal medicine PGY-1s at the University of Arizona Affiliated Hospitals, eight family practice PGY-1s at the same hospitals, and six internal medicine PGY-1s at the Tucson Hospitals Medical Education Program. Each PGY-1 was randomly assigned to examine at least one PI between the third and eighth month of his training year (due to scheduling difficulties, every PGY-1 did not have the opportunity to examine both a pulmonary and a cardiovascular PI). The houseofficer was asked to perform the specific specialty examination on the PI and to be prepared to record his findings at the completion of the examination. The PGY-1 was told that his examination of the PI would serve to identify strengths and deficiencies in physical diagnosis skills. Poor performance on the examination would not prevent his progress in the training program.

At the completion of the houseofficer's examination, the PI evaluated thoroughness and proficiency of the examination using a performance checklist and assigned a performance score. Concurrently the houseofficer identified and described the PI's findings on a content checklist. The PI then assumed the role of teacher and reviewed with the houseofficer his examination techniques and his identification and description of findings. The PGY-1 was given an opportunity to review and repeat relevant portions of the examination. The PI provided informal feedback to the houseofficer on his interpersonal skills within the examination setting and suggested methods to increase a patient's cooperation and comfort. The PIs functioned independently of direct physician input; at no time during the 90-minute evaluation session did a faculty physician enter the room.

RESULTS

Correlations between Scores: Correlations between performance and content scores within each specialty area were computed. The correlation coefficient was .05 for the cardiovascular examination and .22 for the pulmonary examination. Neither approached significance. In addition, the correlations between cardiovascular and pulmonary content scores and the cardiovascular and pulmonary performance scores were computed. The correlation between the two content scores was -.22 (not significant). The correlation between the two performance scores was .40 which is significant only at the .08 level with a two-tailed test of significance. These findings were similar to those obtained with the medical students (5).

Correlation of Scores with Program Directors' Evaluation: Six months into the PGY-1 year, two directors of the university-based internal medicine residency program ranked their 12 PGY-1s on the basis of ability and skill. The directors were instructed to use whatever evaluation means they had at their disposal, including subjective impressions from daily rounds, conference presentations, and evaluations received from other attendings. The most intensive personal contact the directors had with the PGY-1s was when one of them was ward attending. Other interactions occurred during weekly professor's rounds and conferences. The correlations between directors' rankings and the content scores obtained by the PGY-1s in both specialties were not significant. The directors' rankings, however, were positively correlated with both cardiovascular and pulmonary performance scores (Table 1). The reliability of the directors' average rankings was .47 when corrected by the Spearman-Brown formula.

Table 1

Correlations between Program Directors' Rankings and Specialty Examination Scores of PGY-1s

	Content	Performance
Cardiovascular	.13	.60*
Pulmonary	.30	.71**

* p=.04 (one-tailed); .08 (two-tailed)
** p=.02 (one-tailed); .04 (two-tailed)

Correlation of Scores with Peer Group Evaluation: In the twelfth month of the PGY-1 year, each university-based internal medicine intern was asked to rate his PGY-1 peers on attitude, ability to acquire a data base, ability to identify problems, and overall clinical competence. A five-point rating scale was employed in the evaluation of each of the four characteristics. No anchoring descriptors were provided. The evaluations were anonymous and each houseofficer was told not to rate himself or any other PGY-1 with whom he felt he had insufficient contact.

Composite peer ratings were then correlated with scores obtained on the PI examinations. Significant correlations were found between the peer ratings and the content scores achieved by the PGY-1s on both the pulmonary and cardiovascular examinations (Table 2). Correlations between peer ratings and performance scores were not significant. The reliability of the average peer ratings of the four individual areas rated ranged from .79 to .89 as computed by the method described by Ebel (6).

Table 2

Correlations between Peer Group Ratings and Specialty Examination Scores of 12 University-based PGY-1s

	Content		Performance	
Characteristic Rated	Pulm	CV	Pulm	CV
Data Base Acquisition	.35	.46	.27	.12
Problem Recognition	.44	.60*	-.18	-.05
Attitudes	.47	.42	.01	-.01
Overall Competence	.54*	.58*	.00	.03
Average Rating	.48	.55	.03	.02

(*p=.05)

Correlation of Scores with Acceptance Committee Rankings: Each university-based internal medicine PGY-1 had been ranked by the internship acceptance committee of the department prior to beginning his internship. The committee ranking was largely subjective and based on the dean's letter, individual letters of recommendation, and a personal interview. The dean's letter was weighted most heavily. The relative rankings by the acceptance committee correlated significantly with the content scores on the cardiovascular examination (r=.62; p=.02), with overall peer ratings (r=.52; p=.04), and with peer evaluation of each PGY-1's ability to obtain a data base (r=.56; p=.03).

DISCUSSION

There was no significant correlation between the performance and content scores achieved by a given houseofficer in either specialty examination. Houseofficers who performed the most thorough examinations were not necessarily those who were able to describe the PIs' abnormalities most accurately. This finding supports the contention that a compulsive examiner may not necessarily be able to describe physical findings any more accurately than a less thorough one. In addition, low correlations between the cardiovascular and pulmonary content scores and the cardiovascular and pulmonary performance scores imply that each specialty examination may require separate evaluation. These findings were similar to those obtained during the student program (5).

Of particular interest were the correlations between the scores provided by the PIs and the evaluation methods currently used by directors of the housestaff program. The subjective rankings of the PGY-1s by the program directors correlate with performance scores. Since the directors have contact with the PGY-1s primarily during rounds and conferences, their evaluations may be based largely on the thoroughness of the PGY-1 in an ordered setting rather than on his actual diagnostic abilities. This finding supports data from the ABIM survey which suggested that in 20% of all programs visited there was inadequate contact between attending physicians and houseofficers to allow adequate assessment of clinical skills (2).

Peer ratings were positively correlated with content scores on the PI examinations but were not correlated with performance scores. The day-to-day cooperation of peers in an actual patient-care setting may provide a prime opportunity to evaluate the diagnostic competence of a houseofficer. If so, this method of evaluation is rarely utilized fully by training programs.

The rankings of the PGY-1s by the internship acceptance committee correlated significantly with peer evaluations and with the content scores on the cardiovascular examination. There was no significant correlation between acceptance committee rankings and PI performance scores or ratings by program directors. The authors cannot fully explain why a significant correlation would result from an acceptance system which is largely based on subjective input. Other observers have indicated that the dean's letter may be a valid predictor of performance (7,8) and the relatively greater weight placed upon that letter by this particular committee may explain at least part of the correlation. In any case, it would appear that the information which is currently being evaluated by this acceptance committee is adequate to predict some components of clinical competence.

The advantages of using PIs as patients, teachers, and evaluators in a competency-based clinical evaluation of medical students have been discussed elsewhere (5). The authors assert that PIs can function effectively as integral components in the evaluation of housestaff. Although the assessments obtained from the PIs encompass evaluations of segments of clinical competence available from other sources, the PIs provide numerical objective data not previously obtained from any other single source. Furthermore, PIs can provide each houseofficer with direct, individual input about his abilities so that areas of weakness can be identified and corrected, thus completing a feedback loop which has often been lacking from training program evaluations of clinical competence (2). Analysis of the checklists may identify individual as well as group errors for program directors. Individual houseofficers with deficiencies in performance, data gathering, or interpersonal skills can be detected early in their training and remediation may be recommended.

CONCLUSIONS

Symptomatic non-physician patient instructors have been trained to function independently of physicians as patients, teachers, and evaluators. Such patients with stable, describable findings have provided a new means of objective evaluation of a houseofficer's ability to perform a physical examination and to gather accurate data from that examination.

The authors would not presume to conclude that the current design of the PI program allows for precise prediction of the performance of the houseofficer in a real patient setting. Rather, the program is an attempt to quantify objectively selected critical components of clinical competence. The program continues to be under expansion and development. It is being used for other levels of housestaff training and continuing medical education for practicing physicians. It is hoped that expansion of this method will eventually permit objective evaluation of the total process of the clinical encounter.

REFERENCES

1. Futcher PH, Fleishman R: Evaluation of the Competence of Internists. Trans Am Clin Climatol Assoc 83:167-176, 1971.

2. Futcher PH, Sanderson EV, Tusler PA: Evaluation of Clinical Skills for a Specialty Board During Residency Training. J Med Educ 52:567-577, 1977.

3. Leake HC III, Barnard JD, Christophersen ER: Evaluation of Pediatric Resident Performance During the Well-Child Visit. J Med Educ 53:361-363, 1978.

4. Schuller AB, Stotz R: The Medical Record Review and Resident Education. J Med Educ 52:133-137, 1977.

5. Stillman PL, Ruggill JS, Rutala PJ, Sabers DL: Patient Instructors as Teachers and Evaluators. J Med Educ 55:186-193, 1980.

6. Ebel RL: Estimation of The Reliability of Ratings. Psychometrica 16:407-424, 1951.

7. Tremonti LP: Intern Selection. Ann Intern Med 79:277, 1973.

8. Gordon MJ, Lincoln A: Selecting a Few Residents From Many Applicants: a New Way to be Fair and Efficient. J Med Educ 51:454-460, 1976.

PRECIS

ENHANCING STUDENT/FACULTY ENVIRONMENT INTERACTIONS

MODERATOR: Marilyn Heins, M.D.
University of Arizona, College of Medicine

MEDICAL STUDENT NEEDS: WHAT AND WHEN

Medical school student affairs programming should relate directly to perceived and prioritized student needs at different times during training. This paper describes a questionnaire for providing this information.

AFFECTIVE LEARNING IN MEDICAL EDUCATION

The purpose of this study was to test a theory of affective learning in medical education. Two critical areas of affective learning were isolated and became the focus of the study: (1) Coping with feelings about learning per se, and (2) Coping with feelings about illness, disability, death and dying.

EVALUATION OF A MEDICAL SCHOOL LEARNING ENVIRONMENT

The current learning environment, rated by students and faculty members, was compared to the environment intended by the planners of the medical program. Ratings for all three current environmental dimensions were lower than ratings for the intended, but both strong and weak aspects were identified within each dimension.

THE EFFECTS OF GROUP STUDY SKILLS COUNSELING AND APPLIED RELAXATION ON STUDY BEHAVIORS AND TEST ANXIETY IN MEDICAL AND DENTAL STUDENTS

This paper reports on an investigation of the effectiveness of a three session workshop designed to improve study behavior and to decrease test anxiety in first year medical and dental students. The results indicate that a combined treatment of study skills counseling and applied relaxation significantly enhanced the participants' self-reported study skills and decreased their test anxiety.

Continued. . .

DEVELOPMENT, IMPLEMENTATION AND EVALUATION OF A PROGRAM TO IMPROVE LECTURE AND PRESENTATION SKILLS

This paper describes the development, implementation, and evaluation of a medical school faculty development program designed to improve lecture and presentation skills. Over 200 faculty, associated with 23 basic science and clinical departments have participated in the program which features faculty-consultant review of teaching, based on: 1) concerns elicited in a pre-observation conference; 2) consultant in-person observation of an actual lecture; and 3) a videotape of that same lecture.

168

MEDICAL STUDENT NEEDS: WHAT AND WHEN

Grant D. Miller, M.D., Elizabeth C. Miller, M.D., Owen C. Peck, M.D.

University of Nevada, Reno, School of Medicine

Although awareness of medical student needs and professional knowledge of their developmental tasks are legitimate stimuli for student affairs program development, data from student needs assessment questionnaires could verify and refine these assumptions. This is particularly important given rapid changes in students, faculty, schools, and given budgetary and time constraints.

This common sense approach is associated with anxiety attributable to several nagging questions. Are student needs being accurately identified? Are student needs being adequately met? Are the programs really needed? Is limited program energy spent according to priorities set by students?

To answer these questions, a search for existing "needs assessment" tools was undertaken. A review of the literature was only partially helpful. Several general descriptive studies of attitudes and personality traits were found (1,2). A number of studies explored the interaction between students and environmental stressors. The most important of these were academically related (3-6). Edwards and Zimet's "inventory or problems and concerns" was the most relevant to this work (7).

Subsequently, the Medical Student Needs Questionnaire (MSNQ) was developed to help answer earlier questions surrounding assumed student needs and student programming. The following criteria influenced the questionnaire design. First, the needs and concerns addressed should be those commonly found in the literature; of these, only those within the sphere of influence of the student affairs office should be included. Second, problems questioned should be related to irritants found in the medical school environment, the magnitude assessed through a Likert self-rating scale. Third, the questionnaire should be short and concise, promoting quick and spontaneous responses. Finally, it should directly elicit program ideas.

The purpose of this paper is to report the results of our survey using the MSNQ and discuss their applicability to student affairs programming.

Reprint requests: Grant D. Miller, M.D., Director of Student Affairs, School of Medicine, Manville Bldg., University of Nevada Reno, Reno, Nevada 89507

METHODOLOGY

The twenty-four item, two-part MSNQ was developed by the authors using the criteria mentioned. The face sheet provided instructions and promised confidentiality. It also requested the following biographical data: sex, age, marital status, living arrangement, estimated hours of study per week, number of hours employed, and if counseling (personal, financial, career planning, or other) had been sought during the previous academic year, for what reason.

Item domain was divided into personal concerns (13 items) and academic concerns (11 items). Personal concerns related to physical health, social support systems, housing, lover/spouse and family relationships, finances, emotional health, sex, sleep, recreation, and substance abuse. Academic concerns related to grade competition, relationship to faculty, ability to absorb required knowledge, study skills, study time, and test performance.

For each item in Part A, students were asked to rate themselves on a five-point scale as to the amount of worry it caused them during the past academic year. The scale ranged from "almost never worried me", given a numerical value of one, to "has worried me much of the time", given a numerical value of five. The mid-point, "has worried me occasionally" was assumed to be the average response and given a numerical value of three.

Part B requested students to list items scored above the mid-point, briefly describe incidents which stimulated the worrying, and provide suggestions which might help alleviate the problem in the future for other students.

The MSNQ was administered to all four classes late spring of the 1979-1980 academic year. It was given to freshm n (N=49) and sophomores(N=46) in a classroom setting; it was mailed with a stamped return envelope to juniors (N=49) and seniors (N=36) with a telephone reminder one week following mailing.

RESULTS

The number of students responding from all classes was 125, or 69 percent (N=180). Of the respondents, 58 percent lived with spouse, family or relative; 67 percent of the students were single, 20 percent held jobs, and 25 percent were female.

The number responding from freshman through senior classes, with percentage of the class total in brackets was 45 (92 percent); 43 (93 percent); 23 (47 percent); and 14 (39 percent). The mean age by class was 24, 26, 27, and 28 years.

The average number of hours studied weekly by freshman and sophomores was 25 and 22 respectively, while both juniors and seniors studied an average of 17 hours. No significant relationships were found between number of hours studied and how much students worried about specific questionnaire items (chi-square).

Of students responding to the questionnaire, 26 percent received counseling during the past academic year. Of these, 58 percent were counseled for personal problems (problems other than financial or career). Of all of those counseled, 53 percent were freshman ($p < .05$ by chi-square), 42 percent of females from all classes sought counseling during the past year ($p < .05$ by chi-square).

For those counseled, nine questionnaire items produced significantly higher mean scores. Items causing the greatest difference in amount of worry were: study skills (concentration and retention), performance on tests, housing or living situation, and competition for grades (.01 level of significance). Those counseled also worried more than their classmates about their ability to absorb all required knowledge, lack of leisure time, and ability to cope emotionally (.05 level of significance).

Females worried significantly more than males about their performance on tests (.01 level of significance), but males worried more about the use of drugs and alcohol (.05 level of significance, separate variance estimate only). Single students worried more about the adequacy of their sex life, conflicts with their parents, and the use of drugs and alcohol (.05 level of significance, separate variance estimate only). Single students also worried more about their relationship to lover than married students (.05 level of significance). Finally, students not living with family (spouse, lover, parents, or relatives) worried more about the relationship to their lover, their housing or living situation, and the use of drugs and alcohol.

In considering data from Part A, when 20 percent or more of respondents rated an item "often has worried me" or has worried me much of the time", i.e. 4 or 5 on a five-point scale, the item was defined as a major worry.

Table I shows the rank ordering of questionnaire items causing major worry to medical student respondents as an aggregate and by class. Items causing major worry are rank-ordered according to the percentage of students rating the item as 4 or 5.

Table II further defines the differences in rankings and provides two results. Not only are there the expected significant differences between upper and lower classmen on levels of worry regarding such issues as career plans, but significant differences also appear unexpectedly between freshman and sophomores on other items.

For each questionnaire item in Part A which caused "worry often or much of the time", respondents were asked to indicate specific incidents which stimulated the worrying, and provide suggestions which might help alleviate the problem for others". Perfect compliance with instructions (included those with no self-rating above the mid-point) was found in 57 percent of the respondents; partial compliance (mentioned incidents, but gave no suggestions) was found in 25 percent of respondents. There were no significant differences in level of compliance between class or other variables studied.

In part B, students responded to 23 of 24 items in Part A, reflecting a wide range of differences in what worries students. The most common suggestions were converting to a pass/fail system, creating study skills sessions, providing tutoring, and increasing available loans inabling students to concentrate on their studies. Lack of leisure time, relationship to spouse or lover, and future career plans produced a large number of suggestions, but increasing the number of loans and scholarships appeared to be the most common suggestion.

DISCUSSION

Though the validity of the MSNQ is questionable, the responses make sense for different classes, i.e. lower classmen were concerned about their relationship to classmates and basic science faculty and about acquiring knowledge, while upper classmen were concerned about debts, residencies, and their relationships to clinical faculty. General similarity to responses found in other studies also supports the validity of this instrument.

The reliability of the MSNQ is also in question. Plans have been made to administer it with minor alterations on a yearly basis. This will determine retest reliability. It will also determine if there are definite changes in worries over time as suggested by this study.

It is our impression that medical students down play worries and tend to use a great deal of denial, particularly if experiencing problems. Although we have no way of knowing, this would certainly depress item scoring and make our findings even more salient.

The main strength of the MSNQ lies in assessment with program development as its goal. Items deal with worries and concerns suggested by the literature. When major worries are identified, information is solicited about their origin and what might be done. This grassroots information source, problem definer, and fund of program ideas is a practical basis for student affairs programming which is preventive in nature.

The results from this study focus attention on two groups vulnerable to worry. As highlighted by Table II, freshman tend to worry more than any other class about more things, and frequently use counseling. The second group is female. They tend to use counseling more frequently than males with 42 percent of all females from all classes seeking counseling during the past year. These data will be useful in supporting new and bolstering existing programs. The latter suggests the MSNQ's usefulness in evaluating existing progams as well.

The MSNQ's ultimate value lies with its ability to itemize concerns common to medical students and prioritize specific problem areas for the individual institution. Given everyone's limited resources this tool can help a school maximize its student affairs efforts. These efforts should always encourage the development of physicians who can humanly acknowledge, identify, and constructively handle needs in themselves and their patients.

BIBLIOGRAPHY

1. Beck, Howard, S.; Blanche, Geer, Everett. Boys in White: Student Culture in Medical School. U. of Chicago Press, 1961.
2. McGuire, Frederick, L., Ph.D. "Psycho-Social Studies of Medical Students: A critical Review". J. Med. Educ. 41:424-445, May, 1966.
3. Funkenstein, Daniel, H., M.D. "The Learning and Personal Development of Medical Students and the Recent Changes in Universities and Medical Schools". J. Med. Educ. 43:883-897, Aug., 1968.
4. Adsett, C., Alex, M.D. "Psychological Health of Medical Students in Relation to the Medical Education Process." J. Med. Educ. 43:728-734, June, 1968.
5. Boyle, Blake, P., M.A.; and Coombs, Robert, H., Ph.D. "Personality Profiles Related to Emotional Stress in the Initial Year of Training". J. Med. Educ. 46:882-888, Oct., 1971.
6. Coburn, David, Ph.D.; and Jovaisas, Al, V. "Perceived Sources of Stress Among First-Year Medical Students". J. Med. Educ. 50:589-595, June, 1975.
7. Edwards, Marc, T., M.D.; and Zimet, Carl, N., Ph.D. "Problem and Concerns Among Medical Students". J. Med. Educ. 51:619-625, August, 1976.

TABLE I

RANK ORDER OF QUESTIONNAIRE ITEMS CAUSING MAJOR WORRY TO MEDICAL STUDENTS BY CLASS*

Questionnaire Item	All Classes	Class+ Fresh.	Soph.	Jr.	Sr.
Ability to absorb all required knowledge	1	1	1	1	4
Adequate study time	2	2	6	-+	-
Study skills (concentration retention)	3	4	3	-	-
Relationship to spouse/lover	4	9	-	6	2
Lack of leisure time	5	6	4	7	-
Lack of enough money	6	8	-	8	1
Financial indebtedness	7	7	-	5	3
Performance on tests	8	3	5	-	-
Relating to patients	9	10	2	-	-
Future career plans	10	-	-	2	5
Ability to become a "good" physician	(11)++	-	-	9	7
Physical health	(12)	5	-	-	-
Competition for grades	(15)	-	-	10	-
Experiences at community training sites	(16)	-	-	4	-
Adequacy of sex life	(18)	-	-	-	6
Relationship to clinical faculty	(22)	-	-	3	-

*A major worry is when 20 percent or more of respondents checked "often worried me" (4) or "has worried me much of the time" (5) on a five-point scale.

+Blanks appear where 20 percent of respondents in class did not vote item as a major source of worry.

++Brackets indicate rankings where less than 20 percent of the aggregate checked four or five on a five-point scale.

-162-

TABLE II

SIGNIFICANT DIFFERENCES IN QUESTIONNAIRE ITEM MEAN SCORES BY CLASS (.05 LEVEL OF SIGNIFICANCE)

Questionnaire Item	Class
Ability to absorb all required knowledge	Fresh. > Soph.
Adequate study time	Fresh. > Soph. Fresh. > Sr.
Study skills (concentration, retention)	Fresh. > Soph.
Performance on tests	Fresh. > Soph.
Relating to patients	Fresh&Soph. > Jr.&Sr.*
Future career plans	Jr. > Fresh. Sr. > Fresh. Jr.&Sr. > Fresh.&Soph.*
Physical health	Fresh. > Soph. Fresh.&Soph. > Jr.&Sr.
Sleeping well	Fresh. > Soph.* Fresh. > Jr.
Experiences at community training sites	Jr. > Fresh.
Relationship to classmates	Fresh. > Soph.* Fresh. > Jr.*
Relationship to basic sciences faculty	Fresh. > Soph.*
Relationship to clinical faculty	Jr. > Fresh.

* Level of significance = .01

AFFECTIVE LEARNING IN MEDICAL EDUCATION*

ELSBETH KAHN, Ph.D., Associate Professor, Department of Community and Family Medicine, University of Southern California School of Medicine; SANDRA L. LASS, Ph.D., Assistant Professor of Clinical Medical Education, Department of Medical Education, University of Southern California School of Medicine; RUSSELL HARTLEY, M.A., Analyst, Department of Research in Medical Education, University of Southern California School of Medicine; HELEN KORNREICH, M.D., Associate Professor, University of Southern California School of Medicine and the Children's Hospital of Los Angeles.

The purpose of this study was to test a theory of affective learning in medical education. The conceptual model for this research was the result of the senior author's observation and experience with first-year medical students over a six-year period in an Introduction to Clinical Medicine Course (I.C.M.). The course goals in the first year include helping students to understand patients and their experience with illness and to understand themselves in relation to the patient. It seemed that as students progressed in learning, they passed through phases of behavior which reflected feelings and attitudes about new learning as well as feelings about illness, disability, death and dying, thus affecting their interactions with patients, fellow students and instructors. Observations suggested that there were underlying behaviors and steps which students appeared to work through and which seemed to involve the beginning of professional behavior. For example, many students, after an initial period of confusion (during which they seemed oblivious to the magnitude of the learning tasks at hand), became painfully aware of the newness of learning in I.C.M. In their anxiety they tended to deny the need to learn. As a consequence, they would belittle the learning task, or see no need to learn at all. In time, they tended to become angry, resentful and depressed. They would express resentment about instructors' emphasis on professional behavior with patients, declaring their preference for their own familiar and more acceptable styles, until gradually they became more open to looking at new ways of thinking and doing. Simultaneously, they would be making adaptations to being around ill persons. Many students were not prepared for what they saw, heard and smelled, or for their own reactions of fear, anxiety, guilt and disgust. Some would patient's bedside. There were expressions of anger and signs of depression as they were coping with the impact of illness. Thus students might be critical of care provided, question negative prognoses excessively or remain unaware of a patient's upset. This behavior, in time, would subside and students would begin to feel at ease around the very sick and disabled and seem to get on better with their learning tasks. Thus it appeared that this might reflect progress through affective learning tasks which have not been previously specified. Consequently, this research was designed to test a theoretical model of affective learning in I.C.M.

Two critical areas of affective learning were isolated and became the focus of the study: (1) Coping with feelings about learning per se, which refers to attitudes about being an undergraduate student all over again, recognizing what

*This was a research project funded by the National Fund for Medical Education from July 1977 to June 1979.

For reprints write: Elsbeth Kahn, Ph.D., Dept. of Community & Fam. Medicine, USC School of Medicine, 2025 Zonal Ave., Los Angeles, California 90033.

lies ahead, and ultimately giving up dysfunctional learning of past education (for example, interviewing style); and (2) Coping with feelings about illness and disability, death and dying, which refers to dealing with the ever present fact and threat of illness and disability (which few students anticipate), and giving up unrealistic ideas (for example, expectations about prognosis). This learning process is seen as central to the integration of professional behavior into the self-concept, lest unresolved feelings affect the objectivity required to care for sick individuals humanely and effectively.

Two basic assumptions are made. First, these are coping tasks in affective development which are inherent in specialized learning situations, e.g., the clinical aspect of medical education. Thus these critical areas of affective learning involve dimensions different from affective learning related to specific course content. Second, progress in these areas of affective learning will influence medical students' progress in clinical learning.

Behaviors referred to above, which students seem to work through, were reminiscent of the "mourning process" described by Lindeman(1) and Kuebler-Ross(2). Further, it appeared that the "mourning process" reflects attitudinal or affective learning as described and defined by Krathwohl, et al.(3). The implications of affective learning behavior thus far described can be conceptualized further as follows: Medical education requires the acquisition of attitudes functional for learning of the professional curriculum, as well as changing and/or modifying attitudes about illness and disability, thus ultimately requiring change in self and therefore change in self-concept.

The task of modifying aspects of self-concept evokes the mourning reaction. The latter is expressed in feelings of loss and grief as the process of relinquishment of aspects of non-professional self, which become dysfunctional to the developing professional self, proceeds.

Extensive search of literature did not reveal investigations of affective learning as described in this study. In recent years articles and books have been published which deal with stresses in medical education and teaching of sensitive areas such as death and dying. Boles(4) discusses the lack of achievement of affective learning goals in an article reporting on a study of freshman medical students interviewing patients.

THE STUDY MODEL

Area I: Learning Per Se. Prerequisite to acquiring professional knowledge and skills is the emotional acceptance of the need for such learning, and for the relinquishment of "lay" knowledge, attitudes and skills. Thus, how the student perceives relationships, style of communication, and roles are all part of self-concept. In the process of learning to work with patients, students find that habitual roles and style are no longer acceptable or functional, and must be modified. Students' emotional reactions are as to a loss of part of self-concept. The process of resolution begins when students can give up non-functional aspects of self and can accept the need to learn new knowledge, attitudes and skills.

Area II: Illness. The task for this Area of learning is the relinquishment of fantasies of "perfect health" and immortality in preparation to coping with the reality of illness. In order to care for sick individuals students must deal with their own vulnerability and mortality. Through fantasies of loss of health, etc., students experience their finiteness. The resulting perception of loss in self-concept gives rise to another loss and grief reaction. The "mourning process" again involves giving up an aspect of self-concept and accepting a new self-concept which will free the student to begin to respond more realistically to perceive threats of illness.

The model incorporates four Phases of affective learning through which students progress, at differing rates in the two Areas, as they cope with the affective learning tasks described. These are Phase 1--Confusion and bewilderment; Phase 2--Denial of the loss; Phase 3--Anger and depression; Phase 4--Reintegration by coping through learning. Learning occurs through coping with change. It appeared that observations could be tested in a systematic approach of study. The model was constructed with an integrated process for each of the two affective learning Areas, and with each phase having behaviors and educational resolutions.

Three major hypotheses are presented. Hypothesis 1 stated that there will be a difference in Progress Score from Time Period 1 (September to November) to Time Period 4 (April to June). Hypothesis 2 predicted that there will be a steady, positive trend in Progress Score through the four time periods. Hypothesis 3 stated that there will be no main effect due to group.

METHODOLOGY

In the I.C.M. course the class is divided into approximately 24 groups, each with a physician instructor and each assigned to one of the School's teaching hospitals. Typically, students spend about half the class period seeing patients and the remainder in small group discussion. The study sample consisted of six randomly selected I.C.M. groups of six students each, or a total of 35 students (1 dropout) in the first year of the study, and of five randomly and one non-randomly chosen group, or a total of 35 students (1 dropout), in the second year of the research. Participation in the project was on a voluntary basis. Only one group out of the original random sample decided not to participate. Students and instructors were oriented to the project without being given information about the conceptual model.

The data gathering instrument, which was developed and used by the observers for documenting behaviors, consisted of 80 possible behaviors for the first year and 62 behaviors for the second year of the study. The behaviors were distributed over the four Phases of each Area of learning. They were behaviors which the senior author had actually observed over several years of teaching in I.C.M. Since the integrative process for both Areas was the same, discriminating between behaviors for each Area was difficult and remains one of the limitations of this study, as was the length and organization of the instrument.

For purposes of scoring, observation sessions were grouped into Time Periods to smooth over variations in student activity, chance variations in observer attention from session to session, and deviations from the scheduled observation days. These groupings were: Period 1--September to November; Period 2--December to January; Period 3--February to March; and Period 4--April to June.

Progress Score was defined as the level which a student had achieved in progress through the four Phases at the end of each Time Period. It was determined by weighting the number of observations in each phase at one observation period by the number of the Phase (1-4) and dividing by the total observations in all Phases during one observation period.

Trained observers collected data during the discussion periods described above. Observer reliability checks were performed each year prior to the beginning of the data gathering process. During the first Project year each group was observed nine times, and during the second year each group was observed twelve times.

RESULTS

Results pertain to the second year of the study. The first year was considered to be a pilot on the basis of which adjustments were made in the observation instrument and in the methodology for the second year of the study.

Hypothesis 1: There will be a difference in Progress Score from Time Period 1 (September to November) to Time Period 4 (April to June). An Analysis of Variance Repeated Measures Design demonstrated that there was a significant difference in Progress Score over time in Area I and Area II scores ($p < .0001$). This difference held in both Areas of learning ($p < .0001$). This hypothesis was accepted.

Hypothesis 2, which predicted that there would be a steady, positive trend in Progress Score through the four Time Periods within each Area, was accepted ($p < .0001$).

Hypothesis 3: There will be no main effect due to group. This hypothesis was rejected. The analysis of interaction between Area scores and group revealed an interaction effect between Area scores and groups as well as a main effect due to group ($p < .001$).

DISCUSSION

The positive finding for Hypothesis 1 supports the theory that there is a process of learning which takes place in the affective domain. With reference to Hypothesis 2, Tables 1 and 2 present the group means for each Time Period for Area I and Area II. Since there was a steady, positive trend within each Area, the authors believe that it is possible for two independent areas of learning to exist as suggested by the study theory. This assumption will be the subject of further study.

Inspection of Tables 1 and 2 reveals an upward trend for all groups while Figure 1 shows the upward trend of Progress Scores through the Time Periods. The results suggest that students give behavioral responses characteristic of each of the four integrated Phases described in the model and that their learning patterns followed the conceptual framework. This may mean that students work through four Phases of a loss and grief experience related to changing self-concept, and that beginning professionalization can be observed in the fourth Phase. Thus, there was progress as predicted in affective learning. While it was assumed originally that there would be no main effect due to group, group process is a highly significant factor which might have had an effect upon progress in affective learning in the two Areas under study.

CONCLUSIONS AND IMPLICATIONS

Data were analyzed by: (1) process in affective learning--signifying changing and reintegration of self-concept, and (2) areas of learning--Learning Per Se and Illness and Disability. The data analysis related to Hypothesis 1 and 2 appears to support the theory of process in affective learning in clinical experience. The authors are cognizant that affective learning takes place in other areas of the curriculum as well. It is possible that some behaviors observed related to affective learning outside the immediate clinical situation. The difference in trends in Area Scores suggests, however, that a unique aspect of learning process may have been tapped by this study. Consequently, even if affected by other variables, the difference in trend in Area Scores suggest support for a theory of two specific areas of affective learning in medical education as described. Further, the theory may be summarized as follows: observable behavior suggests that affective learning requires (1) giving up old beliefs and behaviors; (2) changing self-concept; and (3) working through a mourning reaction. We believe that these changes are measurable in the clinical setting where the behavior is manifested.

The effect of group differences needs to be the subject of further study. Different students progress at differing rates. Different groups have internal characteristics which determine behavior of group members, in this instance, progress through Phases of affective learning. Class characteristics and therefore group characteristics vary from year to year. As Becker, et al.(5) determined in their study, medical students as a group develop perspectives and a culture which relates to how they solve the problems which they perceive in relation to various aspects of medical education. Yet, while we speculate that different I.C.M. groups may have had different perspectives with regard to how to cope with learning in I.C.M. specifically, and with the specified affective learning tasks by implication, all groups progressed through Phases 1-4 in the manner predicted.

Our model has implications for curriculum content planning. This is especially true for the mix of cognitive and affective learning tasks, which would allow the student to achieve mastery of affective learning tasks as rapidly and efficiently as possible. In that sense, we see this approach not only as important to the education of the physicians who will be able to practice in truly humanitarian, caring and patient-centered manner, but also to the efficiency and economy of the process of professional education. The authors anticipate I.C.M. instructors exploring the use of a revised version of the observation instrument as part of the evaluation process expanding the experiment to other areas of the curriculum and testing this theory in a longitudinal study of medical education.

REFERENCES

1. Lindeman, Erich: "Symptomatology and Management of Acute Grief," Am. J. of Psychiatry, September 1944, Vol. 101, pp. 144-148.
2. Kuebler-Ross, Elisabeth: On Death and Dying, McMillan Co., 1969.
3. Krathwohl, D.R., et al: Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook II: Affective Domain, New York: David McKay Co., Inc. 1964.
4. Boles, B.Kaye: "Interpersonal Skills, Development in Health Profession Education," J. of Allied Health, 5(4) Fall 1976, pp. 23-30.
5. Becker, H.S., Hughes, E.C., Greer, Blanche and Strauss, Anseln, L.: Boys in White Student Culture in Medical School, The University of Chicago Press, 1963.

TABLE 1

ICM GROUP MEAN AREA I SCORES FOR EACH TIME PERIOD

1978-1979

TIME PERIOD		Groups 1	2	3	4	5	6	ALL
1. Sessions	1-3	1.90	2.07	1.92	1.89	1.53	1.75	1.84
2. Sessions	4-6	2.30	2.18	2.52	2.38	2.11	2.19	2.28
3. Sessions	7-9	2.53	2.64	2.86	2.53	2.65	2.70	2.65
4. Sessions	10-12	3.06	3.31	3.45	2.91	3.18	2.97	3.15

TABLE 2

ICM GROUP MEAN AREA II SCORES FOR EACH TIME PERIOD

TIME PERIOD		Groups 1	2	3	4	5	6	ALL
1. Sessions	1-3	2.10	2.08	2.01	2.50	2.12	2.13	2.16
2. Sessions	4-6	2.49	2.57	2.16	2.62	2.11	2.42	2.39
3. Sessions	7-9	2.94	3.00	2.60	3.15	2.51	3.01	2.86
4. Sessions	10-12	3.19	3.53	3.56	3.13	2.98	3.44	3.30

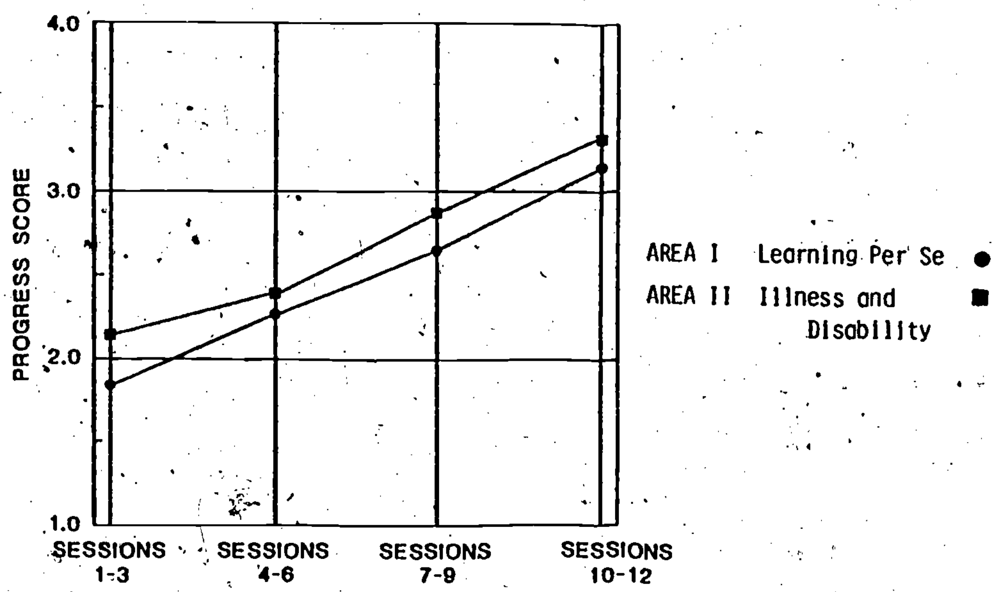


FIGURE I. Program Score vs Time
1978-79 Data

EVALUATION OF A MEDICAL SCHOOL LEARNING ENVIRONMENT

A. Harris, Ph.D., Faculty of Medicine, University of Calgary

This investigation of the learning environment at the University of Calgary medical school included a factor necessary for systematic evaluation--a criterion. The learning environment intended by the planners of the medical program (1968-1970) was the criterion against which the current learning environment was compared. The study asked three questions: How does the current learning environment compare with the environment envisaged by the planners? Do students and faculty members differ in their perceptions of the environment? What strengths and weaknesses, relative to the environment intended by the planners, are present?

Investigators of medical school learning environments have asked various questions. Rothman and Ayoade (1970) investigated the usefulness of their Learning Environment Questionnaire for curriculum evaluation and Levy (1974) asked if that questionnaire was sensitive to the effects of change in students' perceptions. Differences between the perceptions of faculty members and students were investigated by Sheehan (1970). Edwards and Zimet (1976) wished to identify problems and concerns among medical students. Other investigators have studied stress in medical school settings (Marshall, 1978; Arnold, 1978; Royer, 1978). Although the questions differed, the studies were predicated on the important relationships among the psychological and social dimensions of the learning environment. This study was similarly developed.

The three dependent variables for this study were three dimensions reported to be common to several different psychosocial environments (Insel and Moos, 1974). These are Relationship, Personal Development, and System Maintenance and Change. Relationship refers to the nature and intensity of personal relationships within the environment. Student to student and faculty to student interactions belong in that dimension. Students' professional and personal growth characterize the second dimension. The third dimension refers to the orderliness of the environment, the clarity of expectations, means of control, and responsiveness to change.

INSTRUMENT

A questionnaire containing 118 statements based on the philosophy and goals of the Calgary medical program (Cochrane, 1968) was developed for the study. Three faculty members knowledgeable in the philosophy and policies of the medical education program reviewed the questionnaire, decided whether each statement was consistent with the philosophy of the medical school, and identified aspects of the environment not included in the questionnaire. Several revisions were made as a result of their suggestions. Then each statement was placed in one of the three environmental dimensions described previously.

The questionnaire was administered to 104 first- and second-year medical students who were asked to indicate on a five-point scale the extent to which the statement described an aspect of the current learning environment.

The psychometric characteristics of the questionnaire were examined by analyzing the data from the students. Statements were deleted using an iterative procedure described by Nunnally (1967). Over several trials 67 statements were deleted either because some statements correlated to about the same extent with more than one dimension or because some statements did not correlate highly with the dimension in they belonged. In the latter case, inclusion of the statements did not appreciably increase the alpha reliability coefficient for the dimension. The purpose of the analysis was to achieve a relatively homogeneous set of statements within each dimension, but also to obtain relatively heterogeneous dimensions.

After modifications resulting from the analysis described above, the questionnaire consisted of 51 items, 10 of which were in the Relationship Dimention, 17 in the Personal Development Dimension, and 24 in the System Dimension. The internal consistency reliability coefficients (alpha) for the Relationship, Personal Development, and System Dimensions were .84, .83 and .87 respectively.

METHOD

The responses of the 104 first- and second- year medical students to the 51 statements were taken as an estimate of the current learning environment as perceived by the students. There were no significant differences between mean responses on the three dimensions for the first and second-year students, so their data were pooled for analysis.

In addition, 31 full-time faculty members responded to the questionnaire. Their total teaching hours were in the upper quartile for all full- and part-time faculty members' teaching hours. That group was assumed to have more complete information about the learning environment than faculty members with fewer teaching hours. The responses from the 31 faculty members were considered an estimate of the current learning environment, as perceived by the teaching faculty.

To obtain an estimate of the intended learning environment (the criterion for the study), seven faculty members (judges) were asked to sort the 51 statements into five groups, from most important to least important, in terms of the philosophy and program of the medical school. Three of the faculty members had been involved in planning the medical program during the medical school's inaugural period and the other four had arrived shortly thereafter. All seven had been active in policy formation, curriculum development, and teaching since their arrival. The interrater reliability coefficients for the seven judges' ratings (unadjusted means) calculated from analysis of variance, repeated measures, of the ratings (Winer, 1962, pp. 124-132), were .89, .63, and .66 for the Relationship, Personal Development and System Dimensions respectively.

The responses from the three groups were compared to see if the estimates of the current learning environment (represented by the ratings of the students and faculty members) and the intended learning environment (represented by the ratings of the judges) differed.

RESULTS AND DISCUSSION

The mean ratings from the three groups for the three dimensions of the learning environment are given in Table 1. To test for significant differences between the means, the data were subjected to a multivariate analysis of variance, an approach which controls for interdependencies in the data which might otherwise result in spuriously high F ratios.

TABLE 1

Mean Ratings for Three Dimensions of the Learning Environment

		Relationship (10 items)		Personal Development (17 items)		System Maintenance (24 items)	
	N	Mean	sd	Mean	sd	Mean	sd
Intended (judges)	7	43.57	3.10	72.14	5.87	[illegible]3	9.62
Current (students)	104	35.57	6.84	60.38	9.25	[illegible].06	14.36
Current (faculty)	31	34.26	3.19	53.52	7.38	[illegible].19	9.66

The multivariate F ratio given below Table is an overall test of significance for the three groups across the three dimensions. When that F ratio is significant (in this case $P < .0001$) it is appropriate to examine the univariate F ratios for differences between the means for the groups on each dimension, and as Table 2 shows the differences were significant.

TABLE 2

Univariate F ratios for Three Dimensions

	df	Univariate F	p
Relationship	2, 139	6.70	.002
Personal Development	2, 139	15.06	.0001
System	2, 139	12.91	.0001

Multivariate F (6,276) = 8.53, $p < .0001$

Multiple paired comparisons were then computed for each pair of means for each Dimension. As Table 3 shows, both students' and faculty members' ratings differed significantly from the judges' ratings for all three dimensions. However, ratings from students and faculty members differed significantly only for the Personal Development Dimension.

The results indicate that the current learning environment does not yet match the intentions of the planners for all three dimensions. On the other hand, the gap between the intended and the current environment was not large. The mean current ratings were proportionately .81, .84 and .73 of the mean intended ratings for Relationship, Personal Development and System dimensions respectively. The intentions of the planners are being implemented to a considerable extent.

TABLE 3

Multiple Paired Comparisons for Judges, Students and Faculty on Three Dimensions

	Relationship Protected t*	p	Personal Protected t	p	System Maintenance Protected t	p
Judges/students	3.35	.01	3.44	.01	5.08	.01
Judges/faculty	2.55	.02	5.08	.001	4.34	.01
Students/faculty	1.05	ns	3.83	.01	.78	ns

* Bock (1975)

Consideration of significant differences only may miss useful information, in particular, the aspects of the environment most or least in need of improvement. A further analysis was done to identify strengths and weaknesses indicated by the differences in mean ratings for items. Because item means are less stable and less reliable than dimension means, only the differences that exceeded the largest item standard deviation or were less than the smallest standard deviation were considered. Twenty items were identified using the foregoing procedure. Six described relatively strong aspects of the learning environment, that is, the differences between the intended and current ratings were very small. Fourteen described relatively weak aspects of the learning environment. In the Relationship Dimension an important strength was the assistance students give each other, an aspect of the environment apparent to most observers. A weakness in the environment related to clarity of expectations. Four statements referring to statements of goals and objectives received low ratings, all of which were in the System Dimension.

Uncertainty about objectives will probably continue to be a problem in medical education for some time yet. The evidence in this study may indicate not a weakness in the program but the present state of knowledge about developing ob jectives. Researchers and educators have not yet developed completely satisfactory models for the task in spite of much effort (Popham, 1975). Nevertheless some uncertainty about goals and objectives may be reduced in the Calgary medical school with the recent implementation of a new approach to stating objectives, one which emphasizes problem solving.

Faculty members' ratings for the Personal Development Dimension were lower than students' ratings, which may be related to the difficulty in clarifying expectations, indicated in the System Dimension. On the other hand, students may be over-rating their accomplishments or faculty members may have more stringent expectations than the students do.

The results of the evaluation of the Calgary medical school's learning environment indicates that some aspects of the environment intended by the planners have come close to realization but others require more attention before they match what was intended. The questionnaire has been a means of obtaining useful information. If a sufficiently sensitive instrument can be developed it could be one means of measuring changes in the learning environment occurring through the coordinated efforts of both faculty and students.

CONCLUSIONS AND IMPLICATIONS

The immediate concern of this investigation of the learning environment at the University of Calgary medical school was to evaluate the current environment against a criterion, namely the environment intended by the planners of the undergraduate medical program. The results of the evaluation are of importance locally. The use of a criterion is important for evaluations of all medical school learning environments, because in the absence of explicit criteria the environment may be judged inaccurately in terms of the instrument scale or by some arbitrary and inexplicit impression of what an appropriate learning environment is.

Characteristics of learning environments vary among medical schools and some common dimensions are needed to help identify advantages and disadvantages of learning environments in relation to different philosophies of medical education. The three environmental dimensions proposed by Insel and Moos (1974) and used in this study warrant inclusion in investigations of other medical school settings to see if the dimensions form a common basis for analyzing learning environments across different settings.

Concern for high standards of medical education and professional competence is evident in the time and effort expended by medical educators on undergraduate curriculum development, instructional strategies, and admission procedures. An important additional component of medical education which has received relatively less attention (at least in the literature on medical education) is the learning environment. Yet accumulating evidence shows that properties of the environment influence behavior (Walberg, 1977). In other words, how students perceive the learning environment can affect their performance in that setting. Therefore, a learning environment should be evaluated in terms of its effect on students' performance. Systematic investigations of medical school learning environments are needed in order to provide information from which to develop intervention studies. Then the important questions about how learning environments can be changed or maintained to have a useful effect on students' performance can be researched.

REFERENCES

ARNOLD, L. (1978) "The learning environment of a six-year program in medical education." Paper presented at the Seventh Annual Conference on Research in Medical Education, New Orleans.

BOCK, R. D. (1975) Multivariate statistical Methods in Behavioral Research. New York: McGraw-Hill.

COCHRANE, W. A. (1968) "Philosophy and program for medical education at the University of Calgary, Faculty of Medicine." Canadian Medical Association Journal, 98:500-505.

EDWARDS, M. T. & ZIMET, C. N. (1976) "Problems and concerns among medical students." Journal of Medical Education, 51:619-625.

INSEL, P. & MOOS, R. M. (1974) "Psychological environments: expanding the scope of human ecology." American Psychologist, 29:179-188.

LEVY, M., MORSE, P. K., LEIBELT, R.A., DALLMAN, J.J., & MCDONALD, T.F. (1973) "Use of the learning environment questionnaire to assess curricular change." Journal of Medical Education, 48:840-845

MARSHALL, R.E. (1978) "Measuring the medical school learning environment." Journal of Medical Education, 53;98-104.

MOOS, R. (1973) " Conceptualizations of human environments." American Psychologist, 28;653-665.

NUNNALLY, J. C. (1967) Psychometric Theory. New York: McGraw-Hill.

POPHAM, W. J. (1975) Educational Evaluation. Englewood Cliffs:Prentice Hall.

ROTHMAN, A. I. & AYOADE, F. (1970) "The development of a learning environment:a questionnaire for use in curriculum evaluation." Journal of Medical Education, 45:754-759.

ROYER, J. A., HUEBNER, L., CORDES, D.L. (1978) "Medical School environmental assessment." Paper presented at the Seventh Annual Conference on Research in Medical Education, New Orleans.

SHEEHAN, T. J. (1970) "Medical school climate: comparative faculty and student views." Journal of Medical Education, 45: 880-887.

WALBERG, H. J. (1976) "Psychology of learning environments." In L. S. Shulman (Ed.) Review of Research in Education. Itasca: AERA.

WINER, B. J. (1962) Statistical Principles in Experimental Design. New York: McGraw-Hill.

187

THE EFFECTS OF GROUP STUDY SKILLS COUNSELING AND APPLIED RELAXATION ON STUDY BEHAVIORS AND TEST ANXIETY IN MEDICAL AND DENTAL STUDENTS

David G. Schroeder, Ph.D.*
University of Connecticut, School of Medicine

Purpose of the Study

This study investigated the efficacy of a program designed to enhance study behaviors and to decrease the debilitating effects of test anxiety in first year medical and dental students. The research goal was to assess the effectiveness of study skills training and applied relaxation as active coping skills.

Background and Theoretical Framework

Academic stress and examination anxiety frequently have been reported as primary concerns of first year medical and dental students (Coburn & Jovaisas, 1975; Fredricks & Mundy, 1968; Sherlock & Morris, 1972, Edwards & Zimet, 1977). These students have demonstrated the academic ability for the successful study of medicine or dentistry, but some have not developed the higher level study skills necessary to organize, memorize, and hopefully synthesize such a large amount of material in such a short time. When faced with this intensified academic demand the professional school student must often re-evaluate and modify the "process" that has previously been used in learning. Concern about one's performance and the concomitant fear of failure contributes, in turn, to an increase in test anxiety.

Scant attention has been paid to the study skills of students in professional schools. Twenty years ago the question "Study Skills Courses in Medical Schools?" was originally posed. The authors answered their rhetorical question by suggesting that medical students could improve their study skills, that a study skills course designed to address the specific needs of medical education may be beneficial, and that the time was ripe for controlled study (Entwisle & Entwisle, 1960). During the intervening years we observe little controlled research and only an occasional glance toward the role of study skills in either medical or dental education. Shatin (1967) reports that for a group of first year medical students there was no difference between those in the upper and lower grade point average (GPA) quartiles and their total scores on a study habits inventory. However, there was a wide range of inventory scores and many negative study behaviors were reported. Assessing a study skills program for medical students, Holcomb & Brown (1972) found no statistically significant difference in grade point average between an experimental and control group, but the participating students reported a feeling of self-improvement in their academic performance. Weinstein & Gipple (1974) investigated the relationship of study skills

---

*Requests for reprints and appendices should be sent to David G. Schroeder, Ph.D., Assistant Professor, Department of Research in Health Education, Farmington, Connecticut, 06032.

to academic achievement in first and second year medical students. They found no significant difference between first and second year classes on the total study skills inventory score. However, correlations between inventory scores and several premedical and medical school measures indicated that study skills were more highly associated with medical school achievement than to the aptitude measures, especially among freshman. Based upon this correlation between study skills and academic achievement for freshman, the authors suggest that study skills training should occur early in the medical school experience. No controlled research was located investigating the effects of study skills training with dental school students. Aranda & Henry (1975) did highlight the need for developing basic study skills as a component of a comprehensive academic reinforcement program for dental students.

Educational and counseling psychologists have actively investigated and applied the theories of learning to the development of practical methods of study. Conscious, systematic approaches to study have been advocated to facilitate the process of transmitting external information to a learner's memory (Hanau, 1978; Maxwell, 1979; Robinson, 1961). By developing higher level study skills the number of study hours is frequently reduced, because the effective study hours are increased. Study skills training has productively included information about concentration, memory, study systems, reading, time management, taking lecture notes, preparing for, taking, and learning from examinations and coping with test anxiety. The relationship of study skills to the effective treatment of test anxiety has been supported by several researchers who have developed treatment programs that combine relaxation or systematic desensitization with study skills training (Allen, 1971; McCordick, Kaplan, Finn & Smith, 1979; Mitchell & Ng, 1972; Spielberger, Anton & Bedell, 1976).

Test anxiety was originally defined as a learned anxiety drive which interferes with the completion of an evaluative task. Mandler and Sarason originally characterized test anxiousness as "feelings of inadequacy, helplessness, heightened somatic reaction, anticipations of punishment or loss of status and esteem, and implicit attempts at leaving the test situation" (Mandler & Saranson, 1952, p. 166).

A body of research has supported Libert and Morris (1967) who argued that test anxiety consisted of two major factors: worry and emotionality. Each component has been theoretically pursued by both cognitive and behavioristic schools of thought. The worry component or "cognitive concern" about performance is considered by cognitive theorists as the most troublesome. The division of attention between self and task was pursued and investigated by Wine (1971) and Sarason (1972). A synthesis of an integration of this cognitive perspective has been offered by Sarason (1975) who combined the previously elaborated cognitive concepts of direction of attention and negative self-talk that distracts the learner's attention. Saranson's intervention for test anxiety focuses on the modification of self-defeating statements, "I'm going to fail." Meichenbaum, (1972; 1977) offered a cognitive-behavior modification treatment which is designed to address both the worry and the emotionality factors. The worry component is addressed through the modification of negative self-ruminations. The test anxious person becomes aware of self-defeating subvocalizations and their relationship to anxiety. The person is taught to replace these negative self statements with incompatible positive self-statements. Meichenbaum's treatment combines this use of self instructional training with a behavioral component which includes modified systematic desensitization and coping imagery. A review of the literature suggests that systematic desensitization is the most utilized and studied behavioral treatment for test anxiety. Briefly, desensitization is a counterconditioning procedure set forth by Wolpe (1958) who combined deep muscle relaxation (Jacobson, 1938) and the visualization of a stressful anxiety hierarchy. Systematic desensitization has been described as a focused, problem-specific type of intervention that does not generalize across anxiety provoking situations (Meichenbaum, 1972; Suinn, 1968). Barrios & Shigetomi (1979) and Tobias (1979) provide reviews of several coping skills training methods and identify applied relaxation (AR) as an efficient method of coping that may be easily learned and generalized to other anxiety provoking situations. Goldfried (1971) emphasized that while learning systematic desensitization a person acquires relaxation as an active coping skill and stressed that increased attention should be given to generalizing the relaxation skill to encounters with other fear provoking stimuli. Applied relaxation has developed as a specific coping technique that includes relaxation training and the application of self-control relaxation to stressful situations. The effective treatment of test anxiety by applied relaxation has been reported by Chang-Liang & Denney (1976) and Deffenbacher & Snyder (1976). The applied relaxation treatment used in the two studies differed in three significant ways. Each study utilized a four phase procedure. The last two phases which include relaxation training and application training are identical. In the Chang-Liang & Denny (1976) study phase one includes an applied relaxation rationale and the use of an anxiety hierarchy. Deffenbacher & Snyder (1976) in contrast, utilizes discrimination training for the identification of anxiety responses cues in the first phase of their procedure. On the basis of limited data Snyder & Deffenbacher (1977) applied relaxation appears to be as effective and more efficient than systematic desensitization in the treatment of test anxiety.

METHODOLOGY

Subjects

During the first month of the academic year and one month prior to their first interim examination, a three session workshop on "Study Skills and Test Anxiety" was offered to a first year class of medical and dental students. The workshop was offered by the Program in Personal and Professional Development at The University of Connecticut Health Center. In order to investigate the effectiveness of the workshop a Pretest-Posttest Control Group Design was utilized. A total of 27 students volunteered to participate. Due to attrition the treatment group concluded with N=15. Twelve additional students were recruited and assigned to a no-treatment control condition.

Instruments

The participants' study behaviors were assessed via the Study Skills Counseling Evaluation (SSCE) (Demos, 1962). Test anxiety was operationally defined as a participant's score on the Suinn Test Anxiety Behavior Scale (STABS) (Suinn, 1969, 1970). Participants and controls were administered both instruments on the day of the first treatment session and during the week following the last session one month later.

Treatment

Study skills and applied relaxation training (SS & AR). The Study Skills workshop spanned one month and included three sessions each two hours in duration.

The workshop addresses: 1) effective study techniques (e.g., concentration, SQ3R Study method, organizing course material, scheduling time, taking lecture notes, preparing for and taking examinations) and 2) training in applied relaxation following the four-step procedure described by Deffenbacher & Snyder (1976). The workshop was conducted by the author and emphasis was placed on the relationship between examination performance, study skills, and effectively coping with test anxiety.

No treatment control (NTC). The students in the control group received no training in study skills or applied relaxation.

Results

Pre-Post Study Skills Counseling Evaluation (SSCE)

The experimental group reported a posttest group mean difference of 14.73 and standard error of 3.7 (14.73 $\pm$ 3.7) demonstrating a statistically significant overall improvement in learning behaviors compared to the control group (-3.83 $\pm$ 2.52) with a $t=4.18$, $p < .001$. Seventeen of the fifty items on the SSCE have been previously determined as "critical." Again the experimental group (6.13 $\pm$ 1.62) reflected significant improvement over the control group, (-3.08 $\pm$ 1.42), $t=4.28$, $p < .001$.

Pre-Posttest Anxiety

The workshop participants reported significant reductions in test anxiety on the Suinn Test Anxiety Behavior Scale (STABS), while the control group reflected an increase in anxiety from pre to posttesting. The contrast was statistically significant for the experimental group (3.07 $\pm$ 7.65) compared to the control group (-15.67 $\pm$ 4.14) with a $t=2.15$, $p < .05$.

Study Skills and Test Anxiety Workshop Evaluation

A written post workshop evaluation was completed by participants who attended at least two of the three sessions. In question number one, students were asked to rate on a four point scale the value of the three sessions, handouts, and group discussion (Appendix A). Question two inquired "Did you benefit from participation in the workshop? Why?" (Appendix B). The final question asked "Will the workshop affect your future methods of studying? Preparing for and taking examinations? Coping with test anxiety?" (Appendix C). The feedback from the above evaluation indicates that the medical and dental students who participated in the workshop found it to be a valuable experience.

Discussion/Implications

Group study skills training and applied relaxation were found to be significantly more effective than the no treatment control condition in improving scores on the Study Skills Counseling Evaluation (SSCE) and the Suinn Test Anxiety Behavior Scale (STABS). Although participants were not randomly assigned, the pretest means on the study skills instrument were not significantly different. In contrast, there was a difference on the pre-anxiety measure identifying the experimental group as significantly more test anxious. The present report indicates that a short term six hour group treatment may be productively utilized in the psychoeducational counseling of first year professional school students. The application of study skills and test anxiety theory to the domain of medical and dental education could significantly contribute to the amelioration of academic stress.

REFERENCES

Allen, G. J. Effectiveness of study counseling and desensitization in alleviating test anxiety in college students. Journal of Abnormal Psychology, 1971, 77, 282-289.

Aranda, T. and Henry, J. Academic reinforcement program. Journal of Dental Education, 1975, 39, 782-785.

Barrios, B. A. and Shigetomi, C. C. Coping-skills training for the management of anxiety: a critical review. Behavior Therapy, 1979, 10, 491-522.

Brown, W. F. and Holtzman, W. H. Survey of study habits and attitudes, form C. New York: The Psychological Corporation, 1966.

Chang-Liang, R. and Denny, D. R. Applied relaxation as training in self-control. Journal of Counseling Psychology, 1976, 23, 183-189.

Coburn, D. and Jovaisas, A. Perceived sources of stress among first year medical students. Journal of Medical Education, 1975, 50, 589-595.

Deffenbacher, J. L. and Snyder, A. L. Relaxation as self-control in the treatment of test and other anxieties. Psychological Reports, 1976, 39, 379-385.

Demos, G. The study skills counseling evaluation. Los Angeles: Western Psychological Services, 1962.

Edwards, M. and Zimet C. Problems and concerns among medical students. Journal of Medical Education, 1976, 51, 619-625.

Entwisle, G. and Entwisle, D. Study skills course in medical school? Journal of Medical Education, 1960, 35, 843-848.

Fredericks, M. and Mundy, P. Dental students: behaviorally observed aspects of professionalization. Journal of American College of Dentists, 1968, 35, 275-281.

Goldfried, M. R. Systematic desensitization as training in self-control. Journal of Consulting and Clinical Psychology, 1971, 37, 228-234.

Hanau, L. The study game, how to play and win. New York: Harper and Row, 1978.

Holcomb, J. D. and Brown S. An assessment of a program of study techniques for medical students. Journal of Medical Education, 1972, 47, 285-286.

Jacobson, E. Progressive relaxation. Chicago: University of Chicago Press, 1938.

Lent, R. W. and Russell, R. K. Treatment of test anxiety by cue-controlled desensitization and study skills training. Journal of Counseling Psychology, 1978, 25, 217-224.

Liebert, R. M. and Morris, L. W. Cognitive and emotional components of test anxiety: a distinction and some initial data. Psychological Reports, 1967, 20, 975-978.

Mandler, G. and Sarason, S. A study of anxiety and learning. Journal of Abnormal and Social Psychology, 1952, 47, 166-173.

Maxwell, M. Improving student learning skills: a comprehensive guide to successful practice and programs for increasing the performance of unprepared students. San Francisco: Jossey-Bass, 1979.

McCordick, S. M., Kaplan, R. M., Finn, M. E., and Smith, S. H. Cognitive behavior modification and modeling for test anxiety. Journal of Consulting and Clinical Psychology, 1979, 17, 419-428.

Meichenbaum, D. Cognitive modification of test anxious college students. Journal of Consulting and Clinical Psychology, 1972, 39, 370-380.

Meichenbaum, D. H. Cognitive-behavior modification. New York: Plenum Press, 1977.

Mitchell, K. R. and Ng, K. T. Effects of group counseling and behavior therapy on the academic achievement of text-anxious students. Journal of Counseling Psychology, 1972, 19, 491-497.

Robinson, F. P. Effective study. New York: Harper and Row, 1961.

Sarason, I. G. Experimental approaches to text anxiety: attention and the uses of information. In C. D. Spielberger (ed.), Anxiety: current trends in theory and research (vol. 2). New York: Academic Press, 1972.

Sarason, I. G. Text anxiety, attention, and general problems of anxiety. In C. D. Speilberger and I. G. Sarason (eds.), Stress and anxiety (vol. 1). Washington, DC: Hemisphere/Wiley, 1975.

Shatin, L. Study skills in medical education: a report and analysis. Journal of Medical Education, 1967, 42, 833-840.

Sherlock, B. J. and Morris, R. T. Becoming a dentist. Springfield, IL: Charles Thomas, 1972.

Snyder, A. L. and Deffenbacher, J. L. Comparison of relaxation as self-control and systematic desensitization in the treatment of test anxiety. Journal of Consulting and Clinical Psychology, 1977, 45, 1202-1203.

Spielberger, C. D., Anton, W. D. and Bedell, J. The nature and treatment of test anxiety. In M. Zuckerman and C. D. Spielberger (eds.), Emotions and anxiety: new concepts, methods, and applications. New York: Lea/Wiley, 1976.

Suinn, R. M. The desensitization of text anxiety by group and individual treatment. Behavior Research and Therapy, 1968, 6, 385-387.

The STABS: A measure of test anxiety for behavior therapy: normative data. Behavior Research and Therapy, 1969, 7, 334-339.

The STABS: A measure of test anxiety for behavior therapy. Behavior Research and Therapy, 1970, 7, 303-311.

Tobias, S. Anxiety research in educational psychology. Journal of Educational Psychology, 1979, 71, 583-594.

Weinstein, P. and Gipple, C. The relationship of study skills to achievement in the first two years of medical school. Journal of Medical Education, 1974, 49, 902-905.

Wine, J. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76, 92-104.

Wolpe, J. Psychotherapy by reciprocal inhibition. Stanford, CA: Stanford University Press, 1958.

DEVELOPMENT, IMPLEMENTATION AND EVALUATION OF A PROGRAM TO IMPROVE LECTURE AND PRESENTATION SKILLS

Donald S. Masler, M.D., Patricia R. Austin, M.A., and Ilene B. Harris, Ph.D.
Office of Curriculum Affairs
University of Minnesota Medical School, Minneapolis, Minnesota

Lecturing is a principal modality of instruction in most medical schools. Moreover, components of effective lecturing, such as organizing content, explaining, and using audiovisual aids, are important teaching skills in any medical school teaching situation. Several programs have been reported that focus on improvement of medical faculty lecturing (Foley, 1976; Irby, 1976). This paper describes the development, implementation, and evaluation of a comprehensive faculty development program designed to improve lecture and presentation skills. The program differs from others described in several ways. It is the only such program, to our knowledge, that incorporates all of the following components: workshop; written guidelines; pre-observation conference; feedback session with a consultant, based on in-person observation of an actual lecture and a videotape of that same lecture; and consultant training in a model of consultation based on mutual exploration of concerns and teaching approaches. It is unique in its main emphasis--faculty-consultant discussion of teaching, based on: 1) concerns elicited in a pre-observation conference; 2) consultant in-person observation; and 3) a videotape. It is the only such program implemented school-wide, rather than for faculty of only one course or department. To date over 200 faculty, associated with 23 basic science and clinical departments, have participated in the program.

DEVELOPMENT

Background. Few faculty at the University of Minnesota Medical School have had any formal training in educational theory or methods, a situation which seems to parallel that described nationally in medical education (Jason, 1977; Page et. al., 1975). Yet, needs assessments and responses to simulated teaching problems posed in FORUM, a newsletter for faculty, suggest that many individual medical school faculty members approach teaching with enthusiasm and would welcome opportunities to enhance their educational skills (Harris, 1979). Therefore, efforts were undertaken to systematically design a school-wide faculty development program (Harris, 1979). For several reasons, initial efforts focused on improvement of lecture and presentation skills. Lecturing is a predominant mode of teaching in the first two years, with over 60% of scheduled hours presented in the form of large class lectures. Faculty were accustomed to having their lectures evaluated by students. But, no resources or constructive help were readily available to foster improvement of lecture and presentation skills.

Send reprint requests to:
Donald S. Masler, M.D.
Office of Curriculum Affairs
University of Minnesota Medical School
Box 33 Mayo Memorial Building
420 Delaware Street S.E.
Minneapolis, MN 55455

We wish to acknowledge the consultation and planning efforts of three other current educational consultants--Mark Patridge, M.A., Diana Randall, M.A., and Terry Schultz, M.A.

First, precepts for effective lecturing and presenting were identified by several methods: reviews of the literature on lecturing, explaining, and using audio-visual aids; reviews of students' evaluations of medical school lectures; and discussions with medical school faculty about their views of effective lecturing. The precepts identified through these methods became the bases for recommendations in workshops, written materials, and feedback sessions.

Second, a review of faculty development programs to improve lecturing was conducted. Foley et. al. reported a program offered to faculty in a single clinical department. Faculty privately rated and viewed videotapes of their own lectures and received ratings and written comments from an educational consultant (Foley et. al., 1976). In-person feedback was offered to faculty, but none utilized this opportunity. In addition, two group sessions were available, focusing on guidelines for improvement of lecture skills; however, attendance was low. Irby et. al. also implemented a lecture skills improvement program in one medical school department in conjunction with the presentation of a newly developed course (Irby et. al., 1976). Consultants attended lectures, took notes, and filled out evaluation forms. They then met with lecturers immediately following lecture presentations to elicit concerns and impressions and provide feedback. This program offered immediate feedback; however, without a videotape, instructor-consultant discussion necessarily focused on recollected impressions and notes. Both of these programs were well-received by participating faculty.

The success of these programs was encouraging and suggested a reservoir of receptivity to the principal approachs--videotape review of teaching and consultant feedback. A review of the general faculty development literature led to inclusion of several additional emphases and components. Consultant in-person feedback based on a videotape and observation of an actual lecture was incorporated into the program as a fundamental element, since research reviewed by Peck and Tucker suggests that observation of a videotaped teaching session with another person (consultant or supervisor) is more likely to effect change in teacher behavior than either consultant feedback without videotape or viewing a videotape without a consultant (Peck and Tucker, 1973). A pre-observation conference was also incorporated into the program, following the clinical supervision model developed by Goldhammer (Goldhammer, 1969). The critical concept associated with the clinical supervision model is that feedback and faculty development efforts should be closely linked with individual faculty members' personal concerns about their own teaching so that they are receptive to feedback and become expert self-critics of their own teaching.

PROGRAM COMPONENTS

Initial contact. The program is offered to faculty in conjunction with teaching in a course so that faculty development assistance is linked with curriculum improvement efforts and is offered at a time when faculty are most likely to be receptive. A letter is sent to each departmental or interdepartmental course faculty four to six weeks before the course begins, outlining the rationale and the general plan for the lecture skills development program. It is emphasized that the program is voluntary and confidential. The letter is co-signed by the course director and by the phase (curriculum year) coordinator so as to communicate colleague and administrative endorsement. A lecture skills assessment form and guidelines for lecturing are enclosed with the letter to orient faculty to criteria for effective lecturing and pique their interest in the faculty development program.

Pre-course workshop. If there is sufficient interest, a workshop is offered to all course faculty before the beginning of their lecture course. The purposes of the workshop are: to provide a forum for sharing ideas and concerns; to provide perspectives and guidelines for effective lecturing; and to orient participants to the consultation process. The workshop format includes: 1) a "model" lecture about lecturing; 2) videotaped examples of certain lecture skills; 3) an orientation to criteria for assessment by using a rating instrument to assess a videotaped lecture; 4) small-group discussions about these ratings; and 5) a role-playing session in which a consultant gives feedback to a lecturer.

Pre-observation conference. Each faculty lecturer is contacted personally by one of five randomly assigned consultants. The purposes and components of the program are further clarified and if the lecturer chooses to participate, the lecturer and consultant determine which lecture will be observed and reviewed. The consultant also asks the lecturer to identify concerns about lecturing. These concerns are explored, using a lecture skills assessment form to facilitate productive discussion and identify a focus for the observation and subsequent feedback session. This discussion, although at times brief, is a fundamentally important component of the program, for it establishes a tone of mutual problem exploration. Moreover, feedback is likely to be most useful if it is based on a contract defining the faculty member's concerns (Goldhammer, 1969).

Lecture observation and videotaping. The previously agreed upon lecture is attended by the consultant and simultaneously videotaped. The consultant uses the evaluation instrument to assess the lecturer on a number of criteria related to organization, delivery and use of audio-visual materials. If possible, the consultant records evidence of effectiveness or problems. Typically, the consultant also reviews the videotape before meeting with the instructor, to do a deep analysis of the lecture. Particular attention is devoted to those areas identified by the lecturer as sources of concern.

Post-observation conference. The consultant and faculty member meet subsequent to the observation. These meetings last for approximately 60 to 90 minutes. The consultant conducts the session following a model of consultation that emphasizes mutual problem exploration. Brief excerpts from a typical dialogue illustrate some main elements of this model as well as the type of feedback shared by consultant and lecturer.

The consultant begins by clarifying the purposes of the consultation and establishing an atmosphere of trust. The consultant identifies the instructors' concerns and intentions and provides specific feedback relevant to those concerns. While the consultant may suggest alternative strategies, the emphasis is on mutual identification and exploration of the benefits and disadvantages of alternative strategies.

C: *There are three sources of data we can use to help focus this discussion, your impressions, my impressions, and the videotape. How would you like to begin?*

L: *I'd like to hear your comments, particularly about my use of slides and then see the videotape.*

C: *Overall, your lecture was quite good. You used some techniques with audiovisual aids that were very effective for illustrating your points. For example, you used the light arrow pointer to illustrate the area of abnormality in the chest x-ray of the patient with farmer's lung disease.*

By slowly and completely outlin[illegible]ries of the abnormality and then identifying the features with[illegible]mality, you aided the students' discrimination and helped them [illegible] the diagnosis was made. Let's look at the tape.

L: *Yes, I know the students don't have much exp[illegible]ience at interpreting x-rays, so I take special care to make it clear.*

C: *Otherwise, how do you think your slides worked out during the lecture?*

L: *I'm concerned that the students couldn't [illegible]e some of them.*

C: *I was sitting in the back row during your lecture and most of your slides were easy to see from that distance, but there were some exceptions. And some of the slides on interstitial fibrosis went by pretty quickly.*

L: *Yes. I'm not sure I should include that series at all. The topic is really covered in another lecture. I could probably delete the whole series and save time for other things I'd like to do.*

C: *What would you like to do?*

L: *I'd like to get the students more actively involved, but I'm not sure how to do it.*

C: *What have you considered doing? ...*

In addition to discussion of the specific lecture, the faculty member is encouraged to express feelings about lecturing and teaching, to ask questions, or to steer the discussion to other educational concerns.

Consultant Training. Consultant feedback is critical to the success of the program. Therefore consultants meet regularly to discuss both lecturing and consultation strategies. During these meetings consultants have reviewed videotapes of lectures and compared their assessments. In addition, consultants share effective lecturing strategies they have observed. These ideas serve as one basis for recommendations to lecturers. Also, consultants discuss typical problems in the consultation process, such as dealing with defensive reactions.

EVALUATION

Several types of evaluation data have been collected: 1) data pertaining to the extent of faculty participation; 2) faculty viewpoints elicited with 24 forced-choice Likert-type items and four open-ended items; 3) faculty viewpoints elicited in structured interviews ~~with faculty participating~~ in a follow-up program; and 4) written and verbal feedback from consultants about their participation in the program.

Extent of participation. During the period from May 1978 to March 1980 over 200 faculty from 23 basic and clinical science departments teaching in 15 courses have participated in the program. When workshops have been offered, approximately 25% have attended, a finding consistent with data reported by Foley (1976). Most faculty give one to four lectures in one course once a year. Of those faculty offered follow-up consultations one year later (n=65), to date 40 (62%) faculty have chosen to participate.

Faculty Viewpoints. Of 160 faculty sent evaluation forms to date, 108 (68%) have returned forms. For the most part, faculty have responded very favorably to the program (m=2.30, when 1=Very Strongly Agree, 2=Strongly Agree, and 3=Agree). Typical overall impressions were: *"(the program) rejuvenated me, after many years* *of teaching, to go back to the basic of good pedagogy."* *"The program has improved the climate for teaching in the medical school by focusing on teaching as a valued activity."* Faculty are particularly positive about individual feedback from consultants (m=1.86) and watching themselves on videotape (m=1.91), an interesting finding in light of Foley's report that faculty did not avail themselves of the opportunity for a conference with an educational consultant (Foley, 1976). It is interesting that the consultants were viewed as more helpful than the videotape in identifying strengths (Consultant m=2.02 vs. Videotape m=2.24), increasing comfort with lecturing (Consultant m=2.04 vs. Videotape m=2.38), and suggesting useful approaches (Consultant m=2.02 vs. Videotape m=2.20), while the videotape was viewed as more helpful than the consultants in identifying weaknesses (Videotape m=2.01 vs. Consultant m=2.11). Furthermore, the consultants were viewed as supportive (m=2.03). These findings suggest the value of consultant feedback in reviewing the videotape. A typical comment about consultants was: *"The main value of the program was talking with an expert consultant concerning my lecture. During the feedback session, she was clearly an 'expert', but seemed to draw ideas out of me rather than 'telling' me what to do."* A typical comment about videotaping was: *"The videotape provided feedback about my strong and weak points as a teacher so that I could see myself as others see me. I learned a lot from just watching myself."*

With respect to outcomes, faculty rated the program highest in helping to plan lectures more effectively (m=2.26), think about new approaches to lecturing (m=2.31), and improve delivery skills (m=2.37). The most frequent comments about lecture outcomes were in the areas of lecture organization and preparation, use of strategies to involve students, and increased comfort with lecturing. For example: *"Will improve lecture organization, especially in making appropriate introductions, transitions, and summaries."* *"I'll get students involved more through the use of questions and case problems."* *"Felt much more comfortable lecturing. Thank you!"*

While most evaluations were positive, negative comments have related to the fact that educational consultants have difficulty giving feedback about the content of lectures. One lecturer commented, *"Style and organization are evaluated; content appropriateness (i.e. factual material) is not."* This problem is inherent in the program as presently constituted.

Faculty viewpoints have also been elicited in structured interviews with faculty who have participated in the follow-up consultations one year after their initial consultations. In general, faculty interviewed have viewed the program as having a significant impact on their development as teachers. One instructor commented: *"The net effect was strongly positive. Discussion last year of approaches to lecturing led to a new concept of lecturing--as a situation in which I can use the small group techniques with which I am comfortable, to promote student involvement. This year, I used case questions periodically in the lecture and allowed time for student deliberation and responses."*

<u>Consultant Viewpoints</u>. Consultants generally were positive and enthusiastic about the program and gratified by faculty response. However, participation in the program is time-consuming with each consultation requiring a total of 4-5 hours, including observation, preparation, contacts with faculty and actual feedback.

<u>Discussion and Conclusions</u>. The results of this faculty development program

in lecture and presentation skills demonstrate a large untapped reservoir of faculty interest and concern for improving their teaching abilities which crosses departmental and discipline lines. It appears that individualized feedback on instructors' teaching performance was the crucial element of this program. Faculty receptivity to these approaches has obvious implications for further directions in faculty development programs generally. At Minnesota, a similar program of in-person observation of clinical teaching has already been implemented and enthusiastically received (Patridge et. al., 1979).

An obvious need for future research is to obtain objective evidence of improved instructor performance. Efforts are underway to collect such evidence in structured interviews to elicit instructors' self-reports of change; observations of subsequent lectures to assess changes in actual performance; and scrutiny of students' evaluations of subsequent lectures. The question still remains: What is the relationship between certain lecture strategies and student performance? Research in this area, although fundamentally important, has been problematic in education generally, due in part to the many variables that affect student performance (Rosenshine, 1971). Despite these difficulties, it would appear that the program as implemented has undeniably been beneficial in enhancing faculty awareness of teaching principles, improving teaching performance, and increasing faculty comfort in teaching.

REFERENCES

1. Foley, R., Smilansky, J., Bughman, E. and Sajid, A. "A departmental approach for improving lecture skills of medical teachers," Medical Education, 10: 369-373, 1976.

2. Goldhammer, R. Clinical supervision. New York: Holt, Rinehart and Winston, Inc., 1969.

3. Harris, I.B. Faculty development program of the Curriculum Affairs Office: Progress report. University of Minnesota Medical School: Curriculum Affairs Office, March 1979.

4. Irby, D., DeMers, J., Scher, M. and Matthews, D. "A model for the improvement of medical faculty lecturing," Journal of Medical Education, 51: 403-409, (May) 1976.

5. Jason, H. et. al. Faculty development survey: Final report. Association of American Medical Colleges, Washington, D.C., December 1977.

6. Page, G.G., Foley, R.R., and Pochyly, D.F. "A survey of interests in teacher training of health science faculty," British Journal of Medical Education, 9: 182-187, 1975.

7. Patridge, M., Harris, I. and Petzel, R. Development, implementation, and evaluation of a program to improve clinical teaching. Paper presented at the Eighteenth Annual Conference on Research in Medical Education, Washington, November 1979.

8. Peck, R.F. and Tucker, J.A. "Research on teacher education," in, Travers, R.M.W., ed., Second Handbook of Research on Teaching, American Educational Research Association, Rand McNally and Co., Chicago, 1973. Chapter 30 (pp. 945-947).

9. Rosenshine, B. "Teaching behaviors related to student achievement: A review of research," in Westbury, I. and Bellack, A.A. (eds.). Research into Classroom Processes, New York: Teachers College Press, 1971. (pp. 51-98)

PRECIS

EVALUATING CONTINUING MEDICAL EDUCATION

MODERATOR: Charles F. Johnson, M.D.
East Tennessee State University, College of Medicine

INVESTIGATIONS IN CPR TRAINING

This study presents an initial assessment and six month follow-up of knowledge and skills of registered nurses and physicians who were successful participants in Canadian Heart Foundation approved one-day training programs in cardiopulmonary resuscitation at the basic life support level. Participants' perceptions of their knowledge and skills, a comparison of the two professions and the influence of roles in CPR incidents on retention of knowledge and skills are discussed.

EFFICACY OF TRADITIONAL CONTINUING MEDICAL EDUCATION IN CHANGING PHYSICIAN KNOWLEDGE AND BEHAVIOR IN THE CARE OF PATIENTS WITH ACUTE MYOCARDIAL INFARCTION

In a controlled study, we assessed the effectiveness of a 2 hour traditional CME program in changing the knowledge and behavior of 23 family physicians caring for 123 patients with acute myocardial infarction. Although significant immediate gains in knowledge occurred, these gains were not translated into significant improvements on 33 patient care practices as assessed by a special medical audit instrument.

A MODEL CONTINUING EDUCATIONAL DELIVERY SYSTEM FOR ISOLATED PHYSICIANS IN THE AREA OF PULMONARY MEDICINE: DEVELOPMENT AND EVALUATION

A model CME delivery system was developed for primary care office practitioners (GP, FP) in non-urban underserved areas in California and was experimentally field tested. Strategies used for the solution of problems in recruitment, curriculum development, evaluation and replication are emphasized.

PATIENT CARE APPRAISAL IN THE AMBULATORY SETTING: EFFECTIVENESS AS A CONTINUING MEDICAL EDUCATION TOOL

Patient care appraisal was tested in 16 family physicians' offices for its effectiveness as a continuing medical education tool. The data shows a clearly significant, positive effect for those physicians in the actively involved group, when compared with an equal number of control physicians.

Investigations in CPR Training

David A. Gass, M.D. and Lynn Curry, Ph.D.

Introduction

Cardiopulmonary Resuscitation (CPR) is being promoted as an effective life-saving skill in the early treatment of complications of myocardial infarction and cardiac arrhythmias as well as accident situations such as drowning or electrocution. (Tweed, W.A., 1980) (Tweed, W.A., 1980). Training in CPR knowledge and skills for those who may be in contact with such people at the time of occurrence of their complications or accident is believed to save lives. (Berhard, W.N., 1979), (Lund, I., 1976). Programs to promote this training are being established according to guidelines formulated by the American Heart Association (A.H.A., 1974) and approved by the Canadian Heart Foundation (C.H.F., 1978)

This study presents an assessment of knowledge and skills of groups of registered nurses and physicians who were successful participants in Canadian Heart Foundation approved one day training programs in cardiopulmonary resuscitation at the basic life support level. Participants' perceptions of their knowledge and skills in relation to the actual knowledge and skills are presented both before and after initial training and before and after follow up six months after training. Similarities and differences between the two professions are presented. The influence of roles in actual incidents of CPR, perceptions and retention of knowledge and skills is discussed.

Methods and Procedures

The administration, nursing staff and medical staff of a community hospital which had requested a program in basic life support CPR was approached and agreed to participate in this study. RN's and MD's who successfully completed the training course were followed up six months after training. A total of 12 R.N's and 13 M.D.'s were included in the study. The training program was conducted according to current standards of the Canadian Heart Foundation by certified basic life support CPR instructors.

At the beginning of each course, participants were asked to fill out a participant survey asking name, training, experience with CPR incidents and experience with CPR training. Secondly, all participants were asked to fill out a questionnaire assessing their perception of their current levels of knowledge, skills and performance ability in CPR. They then completed a fifty question multiple choice quiz to assess their knowledge and performed CPR on a Resusci-Annie for one minute. The latter was assessed by scoring the recording strip as well as a visual check-list scored by the instructor.

This project is funded by a grant from the Nova Scotia Heart Foundation.
Requests for reprints should be sent to David Gass, M.D., Division of Continuing Medical Education, Sir Charles Tupper Building, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.

The training course then proceeded. At the conclusion participants completed the questionnaire assessing perception of their knowledge, skills and performance abilities, a 50 item multiple choice knowledge test and then demonstrated one-man cardiopulmonary resuscitation using the recording Resusci-Annie.

At six months participants were contacted and asked to complete the questionnaire assessing their perceptions, a 50 question multiple choice test, perform one-man basic life support CPR and complete the questionnaire again Actual C.P.R. incidents in the community hospital were identified for medical records and from personal interview with C.P.R. teachers anbd course participants. The roles of various individuals and professional groups in C.P.R. incidence was tabulated.

Analysis:

Factors contributing to knowledge test scores and to performance error rate at six months were analysed using regression techniques. Analyses of roles taken in CPR incidents prior to and after training was possible by differentially weighting the possible roles.

Results:

1. Comparison of professional groups yields observed significant differences between R.N.'s and M.D.'s in the six month knowledge scores. This is true allowing for initial pre-training differences in their knowledge scores ($p=.027$), or not allowing for that initial difference ($p=.011$). Nurses averaged five points less on the knowledge test than physicinas at the initial pretraining point, seven points less at post-training and 8.5 points less at six months post training. The regressions allowing for the initial differences are: Nurse's average knowledge test score at 6 months = constant + (0.47) initial knowledge score −3.38

Physician's average knowledge test score at 6 months = constant + (0.47) initial knowledge score +3.38

There were no significant differences observed between professional groups in the Resusci-Annie recorded error rates. The initial difference between R.N.'s and M.D.'s prior to training was not significant ($p=0.31$), nor was the difference at six months. ($p=0.69$).

For both physicians and nurses there are significant decreases in knowledge test scores and significant increases in performance error rates at six months after training. For physicians the training to six month follow up comparison yielded $p < .01$ for knowledge test score decrease and $p=.015$ for error rate increase. For nurses the comparison were $p=.03$ and $p=.08$.

There is no linear relationship for either R.N.'s or M.D.'s at the six month point between perception of knowledge and the actual knowledge test score. P values range from .313 to .888.

An accurate prediction of nurses' six month knowledge test scores is possible if you ask them to rate their perceived C.P.R. skills or their C.P.R. performance after they have been tested for C.P.R. performance. This perceived performance also predicts the nurses' error rates in the tested performance at six months. The equations are:

Nurses' knowledge =
test scores = 63.3 + (4.05) post test perception of skills (p= .036)

Nurses' knowledge
test scores = 65.5 + (3.5) post test perception of skills) p= .035).

Nurses' error rate= (10.5 - 1.63) post test perception performance)^2(p=.037).

For physicians the predictors are somewhat different. None of the perception measures (knowledge, skills or performance),pre or post testing predicts physicians' error rates at six months. Predictions of knowledge test scores at six months are possible by asking physicians for their perceived skills prior to testing or after testing.

Physicians' knowledge
test scores = 65 + (7) pre-test perceived skills (p=.054)

Physicians' knowledge
test scores = 61.4 + (7.64) post test perceived skills ($p < .0005$).

25 incidents of CPR in the hospital were reviewed and the roles taken by physicians and registered nurses recorded. These were grouped as basic life support skills, supervisory and advanced cardiac life support skills and general unskilled assistance, and the frequency of roles performed in each group by each professional group summarized. (Table 2.) Physicians performed basic life support CPR skills much less frequently than nurses. Both groups performed supervisory and advanced cardiac life support skills.

Discussion:

Both R.N.'s and M.D.'s showed significant effects of entry characteristics (previous knowledge and professional background) on their ability to retain the knowledge component of the standard Basic Life Support C.P.R. course. The longer professional training with its emphasis on pathophysiology and diagnosis may explain the higher scores of physicians on this component. However, no significant differences were seen in the retention of the psychomotor skills learned. Previous indication of the importance of practice with feedback during training programs for lay and professional people in the acquisition of such skills would suggest that this would be the major determinant of retention of such skills and not characteristics of previous professional training or knowledge. (Vanderschmidt, H. 1973) (Berkebile, P., 1973) (Vanderschmidt, H., 1975) (Ammons, R.B., 1958) Comparisons of lay and paramedical personnel following training session with and without practice indicate that while professional training enhanced performance of C.P.R. slightly, the magnitude of this effect was much less than that related to practice opportunities. (Winchell, S.W., 1966). Success in retesting at six months was shown to increase when participants had reviewed course material. (Weaver,F.J., 1979) This positive effect was enhanced by actual practice in the interim. (Weaver, F.J., 1979).

Training lay and health professional groups in basic life support C.P.R. will require a massive commitment of resources. Not the least of these will be the opportunity costs borne by participants. Variations in the basic presentation of the knowledge components should be introduced which take account of the different entry characteristics of participants such as previous knowledge and professional training. Opportunity for review and practice are necessary parts of initial training programs. Retraining needs to be flexible in the light of varying amounts of review and practice undertaken by participants.

Perception of educational need is a powerful motivating factor in voluntary training programs such as these. (Knowles, M., 1970.) In this study, participants' perception of their knowledge did not predict their knowledge test scores although both RN and MD groups perceptions of their skills did predict knowledge. Physicians' performance errors at six months were not predicted by any of their perceptions while nurses' perceptions of their performance abilities did predict performance errors. Practice and/or feedback would be expected to enhance self perception of abilities. (Weinberg, 1977). Registered nurses and physicians have been shown to function in a variety of roles during actual C.P.R. incidents including those of resuscitator, circulator, medicator and recorder (Russo, R.M., 1977). The skills applied by physicians in actual C.P.R. incidents in this study were those of supervision and decision making utilizing their background of knowledge and not actual practice of psychomotor skills. Thus the skills which they have an opportunity to utilize predict their knowledge. R.N.'s employed psychomotor skills frequently in C.P.R. incidents as well as acting in roles of medicator, recorder and circulator. This would enhance feedback on both their knowledge and their skills related to C.P.R. Thus their perceptions of skill predicts knowledge, and both skill and performance ability perceptions predict performance error. The lack of opportunity to perform C.P.R. skills may explain why M.D.'s had a signigicant increase in their performance errors while R.N.'s did not. Continuing opportunities for practice and feedback could be expected to increase the accuracy of perceptions as well as improve the actual performance in C.P.R. incidents and therefore should be an integral part of comprehensive basic life support C.P.R. programs.

TABLE 1.

	Pretraining	Immediate Post Training	6 months Post Training
C.P.R. Knowledge Test Score			
M.D.'s	78.5%	91%	85.5%
R.N.'s	73.5%	84%	77 . %
C.P.R. Performance Errors			
M.D.'s	52	5	34
R.N.'s	68	13	34

TABLE 2

	Physicians	Registered nurses
Basic Life Support CPR skills	7	72
Supervisory and Advanced Cardiac Life Support Skills	55	36
Non-professional Assistance	0	7

REFERENCES

1. Standards for cardiopulmonary resuscitation and emergency cardiac care, American Heart Association and National Academy of Science - National Research Council. JAMA 227 (SUPP.): 834, 1974.

2. Ammons, R.B., Farr, R.G., Block, E., Neumann, E., Dey, M., Marion, R., Ammons, C.H. Long term reterntion of perceptual motor skills. J.Exp. Psych. 55: 318, 1958.

3. Bernhard, W.N., Turndorf, H., Cottrell, J.E., Vea, F., Basak, A.: Impact of cardiopulmonary resuscitation training on resuscitation. Critical Care Medicine 7: 257, 1979.

4. Berkebile, P., Benson, D., Ersory, C., Barnhill,S.,Safar, P.: Public education in heart-lung resuscitation: evaluation of three self-training methods in teenagers in Proceedings of the National Conference on Standards for Cardiopulmonary Resuscitation and Emergency Cardiac Care, National Academy of Sciences, Washington, D.C., May, 1973.

5. Knowles, M.s. The Modern Practice of Adult Education. Association Press, N.Y., 1970.

6. Lund, I., Skulberg, A.: Cardiopulmonary resuscitation by lay people: Lancet ii: 702, 1976.

7. Russo, R.M., Laude, T.A., Sheth, K.A., Howell,A.C., Benjamin, E.C., Gururaj, V.J., Rajkumar, S.V.: Cardiopulmonary resuscitation: assessing emergency department staff performance. N.Y. State Journal of Medicine 1251, 1977.

8. Tweed, W.A., Bristow, G., Donen, N.,: Resuscitation from cardiac arrest: assessment of a system providing only basic life support outside of hospital. CMAJ. 122: 297, 1980.

9. Tweed, W.A., Bristow, G., Donen, N., Kirk, B.W., Evaluation of hospital-based cardiac resuscitation, 1973-77. CMAJ 122: 301, 1980.

10. Vanderschmidt, H., Burnap, T.K., Thwaites, J.K.: Evaluation of a cardiopulmonary resuscitation course for secondary schools. Med. Care 13:763, 1975.

11. Vanderschmidt, H.F., Burnap. T.K.: New aspects of CPR training in Proceedings of the National Conference for Cardiopulmonary Resuscitation and Emergency Cardiac Care, National Academy of Sciences, Washington, D.C., May, 1973.

12. Weaver, F.J., Ramierez, A.G., Dorfman, S.B., Raizner, A.D.: Trainees retention of cardiopulmonary resuscitation. JAMA. 241:901, 1979.

13. Weinberg, A.D., Ullian, M.S., Christiansen, C.H., Raizner, A.: Perceived ability versus actual ability: a problem for continuing medical education in the Proceedings of the Sixteenth Annual Conference on Research in Medical Education, Association of American Colleges, Washington, D.C., Nov. 7, 1977.

14. Winchell, S.W., Safar, P., Teaching and testing lay and paramedical personnel in cardiopulmonary resuscitation. Anaesthesia and Analgesia 45: 441, 1966.

EFFICACY OF TRADITIONAL CONTINUING MEDICAL EDUCATION IN CHANGING PHYSICIAN KNOWLEDGE AND BEHAVIOR IN THE CARE OF PATIENTS WITH ACUTE MYOCARDIAL INFARCTION

Authors: Carl W. White, Janet L. Roseman, Mark A. Albanese, Donald D. Brown, Marcia K. Whitney, Richard M. Caplan - University of Iowa, College of Medicine, Iowa City, IA

Precis: In a controlled study, we assessed the effectiveness of a 2 hour traditional CME program in changing the knowledge and behavior of 23 family physicians caring for 123 patients with acute myocardial infarction. Although significant immediate gains in knowledge occurred, these gains were not translated into significant improvements on 33 patient care practices as assessed by a special medical audit instrument.

In recent years many authors have criticized the traditional forms of continuing medical education (CME) (1 and 2). The formal lecture course conducted at a site remote from the learners practice has been labelled as "unstimulating" and "inconvenient" (3). Topics selected by University faculty specialists having little or no contact with local problems are often felt to be "irrelevant" (3). Learners at these coursesare characterized as merely passive recipients of wisdom. And, despite some positive results (4) traditional CME has been said to be ineffective in improving physician performance (5). These criticisms have led many medical educators to recommend a move away from the large, formal, university-based course in favor of locally developed, small group instruction. The track record of the "new CME" lends support to its enthusiasts. Investigators, often using some form of Medical Care Evaluation, (audit) have demonstrated performance gains for home-based CME (6 and 7). Most of these evaluations, however, are not broadly based but rather concentrate upon a few narrowly focused objectives.

Why then, do many providers of CME persist in continuing to offer traditional, formal lecture courses despite little evidence of efficiency? One reason is that many physicians seem to like them. In a recent survey of primary care physicians in Pennsylvania, Mansfield (8) found that the traditional lecture format was preferred across all age groups and specialties. The same study cites surveys conducted in Kansas, California, Scotland and New Zealand that report similar results. Even with little proof of effectiveness, physicians continue to attend traditional CME courses in large numbers, and give such courses high marks according to "happiness" questionnaires that are often the only method of course evaluation (5).

Thus the controversy over whether traditional CME can be effective continues. Its resolution would appear to depend upon carefully planned courses and indepth evaluation efforts. Levine (9) reports that "rarely does one encounter a systematic and rigorous evaluation study of continuing medical education" and, therefore, there is little "hard data" to support or refute traditional CME approaches. Efforts to rigorously evaluate traditional CME must be made before critics advice can be heeded or dismissed.

The present study focuses on the problems encountered in evaluating performance and knowledge gains resulting from traditional CME. Specifically, the study is designed to: 1) Evaluate course effectiveness in producing knowledge and performance changes; 2) Examine the relationship between knowledge and performance; 3) Evaluate the usefulness of hospital record audit for measuring performance change.

Reprint requests: Carl W. White, M.D., Department of Medicine, University of Iowa, Iowa City, IA 52242.

Because the problems addressed in traditional courses are not locally identified, it is likely that learners will come with widely divergent self-perceptions of deficiencies. This makes the demonstration of systematic gain very difficult. The task, then, is to develop measures that are sensitive enough to detect subtle group changes that may, in fact, represent meaningful and substantial change for an individual.

METHODS A. Course

A carefully prepared 2 hour CME program on the management of patients with suspected or proven acute myocardial infarction (AMI) was presented during the annual Refresher Course for the Family Physician at the University of Iowa. Approximately 175 physicians attended the course. In anticipation of the audit study, precise educational objectives were developed. For each objective, specific patient management procedures were identified and emphasized during the program. The educational objectives were chosen by cardiovascular faculty members to relate to specific areas of therapy that 1) represent important principles of management; 2) are common (high frequency); and, 3) are non-controversial, i.e., there is good agreement among specialists. The information presented was not new, but basically review. The intent was to cover areas of importance, not necessarily anticipating areas of possible physician deficiency. Two faculty members chosen for their educational expertise (one, a medical undergraduate teacher of the year; the other, an American Heart Association Teaching Scholar) presented the material which was delivered in a traditional lecture format.

B. Evaluation

Two 41 item content-parallel forms of a test of knowledge over the program content were developed in the following manner: 1) using the previously identified educational objectives, a pool of test items sufficient to create two content-parallel forms were written by the program faculty; 2) these items were pilot-tested on a group of family practice physicians attending a CME program; 3) item revisions were made on the basis of pilot-test information; and, 4) items were randomly assigned to the two forms of the test. These finalized test instruments were utilized for pre- and post-tests of immediate knowledge. In addition, a third test containing questions about cardiology topics NOT covered in the educational presentation was used to exclude pre-test sensitization. The three test forms were randomly distributed to the participants immediately before and following the educational program. Completed and returned pre- and post-tests were obtained from 127 physicians.

Following the six month post-course period all Refresher Course registrants were sent a letter briefly explaining the audit portion of the course evaluation. Volunteers were requested from among those who attended the AMI segment of the course. Twenty-seven physicians agreed to participate in the study. Two physicians were subsequently not audited because their hospitals were located more than 300 miles from the auditor. Two physicians proved to have no applicable cases. The remaining 23 physicians were located in 21 communities ranging in size from 2,000 to 120,000.

The audit instrument was developed to measure the patient management practice portion of the course objectives and was designed to parallel the test items assessing knowledge of these same objectives. The instrument was developed by the program faculty in conjunction with the Medical Care Evaluation Manager of the Iowa Foundation for Medical Care. This organization, the PSRO instrument for Iowa, has responsibility for developing all of the medical audits for the PSRO designated hospitals in the state. Although the audit instrument was developed following usual procedures and guidelines, the exceptions and instructions for data retrieval were constructed in a manner to allow a specially trained and certified medical abstractor to retrieve maximal amounts of information.

The audit instrument was composed of 80 items. Thirty-three of these represented specific patient care practices (PCP) for which ideal behaviors were specified. The remaining 47 items were patient descriptors designed to ensure careful delineation of the patient population. Possible justifiable exceptions were reviewed by the auditor and the faculty on a case by case basis. Patient care practices (PCP) on the audit were given one of three ratings: 1) yes, the practice was done; 2) no, the practice was not done, and; 3) the auditor could not determine whether or not the practice was done.

The audit covered a six month period before the couse and a six month period after. The same auditor travelled to each physician's local hospital for chart review. Physician and patient identifiers were deleted to insure confidentiality.

Results from the audit were analyzed to give information regarding two major outcomes: 1) changes occurring as a result of the course in individual PCP; 2) changes occurring in the sum of all PCP computed as an index of the general performance of all physicians before and after the course.

RESULTS

Physicians attending all portions of the instructional program on AMI showed significant immediate gains in knowledge. The average score on the two forms of the pre-test was 69.7 ± .84% (n=105) and increased to 82.3 ± 1.4% on the post-test (n=127) ($p < .05$). There was no evidence that sensitization from a content related pre-test increased scores on the post-test. The relatively high test scores on both pre-and post-tests were assumed to confirm the review nature of the content.

The 23 physicians cared for 133 patients with a diagnosis of suspected or proven AMI in the 12 month period under review; 67 were seen in the pre-course period and 66 were seen after the couse. The average number of patients audited per physician over each 6 month period was 2.9 (range 1 to 6).

Of the 33 PCP assessed individually, only 3 showed greater than 10% improvement on the post audit. Mean scores reflecting a composite of all PCP averaged 75.4 ± 4% (SEM) on the audit prior to the instruction, but did not improve following instruction (75.1 ± 4%). Only 4 physicians showed a greater than 10% increase in mean audit scores following the instructional period.

This apparent lack of translation of knowledge gains into significant increases in performance stimulated us to re-analyze the data in an effort to detect possible trends, not reaching statistical significance perhaps because of a small sample size. In order to determine whether a physician's entry-level performance influenced the subsequent results of the instruction, we divided the physicians into low, medium and high performance groups as determined by the pre-course audit.

The analysis of the results was performed twice. The first analysis (all PCP) included all patient care practices. The second analysis (selected PCP) eliminated those patient care practices on which the pre-audit performance exceeded 80%. This analysis was prompted by the concern that if pre-audit performance on certain variables was high (>80%), subsequent gain would be seriously attenuated. The results of these analyses are listed in Table I.

Physician Group	ALL PCP (n=6) Low	(n=4) Medium	(n=8) High	SELECTED PCP < 80% Correct on Pre-Audit (n=6) Low	(n=6) Medium	(n=6) High
Pre-course Audit	.542	.698	.803	.358	.565	.725
Post-course Audit	.636	.700	.818	.415	.657	.697

Table I: Proportion of correct responses on all PCP and selected PCP.

In addition to assessing the mean of the correct responses across all PCP and selected PCP, we were also interested to assess those practices for which a 100% correct response was achieved. We had designated this high criterion as the desired course outcome. The results of these analyses are listed in Table II.

	ALL PCP (n=5) Low	(n=9) Medium	(n=4) High	SELECTED PCP < 80% Correct on Pre-Audit (n=6) Low	(n=6) Medium	(n=6) High
Pre-course Audit	.384	.556	.822	.144	.348	.640
Post-course Audit	.538	.591	.712	.322	.488	.582

Table II: Proportion of PCP correct at a 100% criterion level.

Although no significant differences emerge from these additional analyses, several interesting trends are present. 1) The performance of the lowest group on all analyses tended to increase. 2) Although the mid group's performance was variable depending upon the criterion for analysis, the highest group consistently tended to show a decrease in scores on the audit performed after the course. 3) A high initial performance did not create a ceiling effect and thus mask subsequent gains. 4) Changing the final percentage criterion to 100% correct did not change the major conclusions of the study.

Although the test questions and patient care practice criteria were chosen to reflect the specific course objectives and might be expected therefore to be highly correlated, a careful post analysis revealed that not all of the test items could be matched with a specific "PCP objective". Specific "PCP objectives" could be matched with 4 and 5 test items respectively for the two forms of the test. The correlation coefficient between these matched test and audit items was .542 and .593 on the two pre-test forms, and .716 and .940 on the two post-tests.

Some test items were found to reflect higher order analytical functions reflecting the specific course objectives, but for which an exact audit counterpart item could not be identified. Differences in scores between the audit and test items reflecting different order cognitive functions are exemplified in the following illustration.

OBJECTIVE: To discourage the usage of IM medications to treat the pain of acute infarction.

AUDIT ITEM: % of patients with acute infarction receiving any IM meds for chest pain. Score Initial Audit 66% - Score Final Audit 61%

Question Form A - Patient Jones is experiencing severe crushing substernal chest pain. BP is 90/70. You order Demerol IM. Which of the following might be a consequence of your treatment? (There may be more than one correct answer.)

	% Correct Pre-Test	% Correct Post-Test
* a. The drug may be erratically absorbed	67	86
b. The CPK-MB may be spuriously elevated	71	74
* c. The total serum CPK may be spuriously elevated	71	82
* d. The blood pressure may be altered in an unpredictable manner	62	37
MEAN	68	79

Form B - Which of the following is a preferred method for relieving the pain of acute infarction?

a. Demerol 75 mg IM	83	99
b. Morphine 10 mg IM	92	97
* c. Morphine 8-10 mg IV in incremental (2-4 mg) doses	75	94
d. Demerol 10 mg IV	98	87
MEAN	87	94

* Correct Answers

Although improvements were seen in the average post-test scores on both forms, the audit showed no performance improvement. The disparity between significant increases in knowledge as reflected by increased test scores and no increase in correct behavior as reflected in the audit, was most marked in form B (a lower level cognitive function item). It was less aparrent in form A in which a higher cognitive level test item was used.

DISCUSSION

This study shows that a carefully designed traditional CME course can significantly improve physician knowledge. The translation of such gains in knowledge into improvements in patient care practices was not convincingly demonstrated. However, 3 PCP did show greater than 10% improvement.

We believe that our basic approach is a sound one and that it represents a methodological advance over earlier approaches to the evaluation of traditional CME. The educational program was guided by clearly formulated goals and objectives. Pre- and post-test knowledge gains were assessed with careful attention paid to possible sources of measurement error. Both the knowledge test and the performance test (the audit) were designed to closely parallel the course goals and objectives. But despite the care exercised in the design of the evaluation, methodology issues as well as questions of interpretation remain. 1) What should be the criterion for a "meaningfully improved" PCP? Is a 10% change a realistic goal or should higher levels be required? Should the performance criterion be set with reference to demonstrated knowledge gains? 2) Should a perfect score on the post audit for any single practice be required to signify that a "meaningful improvement" has occurred? Would such a criterion be more likely to show significant results, since small changes in performance may be missed by the auditor? 3) Should only those items be evaluated for which the initial audit score is low and for which significant improvement can easily be demonstrated? 4) Are educational outcomes dependent upon entry level behavior? Would a sample size larger than the current one yield statistically significant differences based on differential entry levels?

The conclusions from this study are based upon 15% of the physicians who attended a 4-day, multi-topic, traditional CME course and who subsequently volunteered to cooperate in the audit study. Whether their behavior is representative of the larger group participating in the educational program is unknown. Since this was a formative evaluation study, we did not attempt to increase the number of physicians who initially volunteered to be audited. We are, however, encouraged to observe a change in attitude of physicians regarding medical audit and a more general willingness to participate in this activity if the goal is to assess educational effectiveness.

The relationship between knowledge measured by specific test items and performance measured by corresponding audit items deserves special attention. Though both our test and audit instruments were designed to reflect the same specific objectives, each test item was not expressly auditable. Some test questions assessed specific behaviors. Others examined understanding of the reasons for performing these specific behaviors. In designing future studies of this kind, special emphasis needs to be placed on specific "PCP objectives" assessed by test performance and precisely correlated to audit items.

Audit of the medical record is an expensive, cumbersome and imperfect method of assessing behavior. Conclusions drawn from the audit are limited to those behaviors which are customarily documented in the record. Many worthy performance objectives of CME can never be assessed in this manner. A comparison of records from multiple institutions, many of small size, makes the task more difficult. The usual audit criteria, acceptable to PSRO and JCAH, generally do not constitute a measurement instrument sensitive enough to detect small but important changes in practice behavior resulting from education. Despite these difficulties, audit of performance is potentially one of our strongest measurement tools. With refinement it should permit us to assess the relationship between knowledge and performance and allow us to determine whether specific patient outcomes have been improved by CME.

Whether any one traditional or locally based non-traditional CME course can significantly effect overall long term behaviours is not certain. The period of time spent in the course, however long, is only a small fraction of total time available for other potentially educational influences: patients, colleagues, mass media, journals, etc. Any new education "dose" is added to many years of previous education and experience, and the effect of the additional increment may not be readily apparent. Notwithstanding these difficulties, sophisticated efforts should continue to be undertaken in an attempt to evaluate the results. The educational method, site, context, audience and many other variables singly and in concert undoubtedly have influences upon behavior whether subtle or profound. Attempts to measure these changes in the most rigorous fashion should continue.

References

1. CE Lewis, RS Hassanein; Continuing medical education - an epidemiologic evaluation. New England Journal of Medicine. 1970, 282 (5).
2. OE Laxdal, PA Jennett, TW Wilson, GM Salisbury; Improving physician performance by continuing medical education. Canad. Med. Ass. J. May 1978, 118.
3. RG Hiss, JK Stross, WK Davis, R MacDonald; Development and evaluation of a community-based pulmonary education system. RIME Proceedings 1979.
4. RM Caplan; Measuring the effectiveness of continuing medical education. Journal of Medical Education. 1973, 48.
5. DA Bertram, PA Brooks-Bertram; The evaluation of continuing medical education: A literature review. Health Education Monographs, 1977, 5 (4).
6. TS Tnui, EL Yourtree, JW Williamson; Improved outcomes in hypertension after physician tutorials, a controlled trial. Annals of Int. Med., 1976, 84.
7. VL Wang, P Terry, BS Flynn, JW Williamson, LW Green, R. Faden; Evaluation of continuing medical education for chronic obstructive pulmonary diseases. Journal of Medical Education. 1979, 54.
8. PK Mansfield, DS Queeney, J VanNieuwkerk, AD Chervenak, CA Lindsay; Attitudes and preferences of Pennsylvania primary care physicians regarding continuing medical education. The Penn. State Univ. Plan. Studies in Cont. Educ. Series 1980.
9. S Levine; Evaluating the evaluators of continuing medical education programs. In Egdahl and Gertman (eds) Quality Health Care, Germantown, MD: Asp. Sys. Cor. 1977.

A MODEL CONTINUING EDUCATIONAL DELIVERY SYSTEM FOR ISOLATED PHYSICIANS IN THE AREA OF PULMONARY MEDICINE: DEVELOPMENT AND EVALUATION

Muriel H. Bagshaw, M.D., Senior Research Scientist
Wylie L. McNabb, Research Scientist
American Institutes for Research, Health Research Group
Palo Alto, California *

Purpose

Keeping physicians informed on new advances in medical knowledge, as well as enhancing retention of skills and knowledge, is a realistic goal for continuing medical education (Williamson, 1977). This report describes the development and evaluation of a model delivery system for an instructional program on office care of asthma for physicians in rural/non-urban practice sites.[1] The approach and methodology utilized in the development of the model is applicable to continuing medical education programs in general, regardless of medical content or practice situation.

Literature Review

Major problem areas in CME for isolated physicians include limited resources, motivation, CME program content, and evaluation.

Limited Resources

Primary care physicians who practice in isolated communities find it difficult to leave practices to attend meetings at distant sites. Non-urban physicians must deal with a broader range of medical problems. Acute medical problems are both more frequent and more severe due to the small size of and distance to hospitals or emergency rooms. Consultants are less available or non-existant. In-migration of recently trained physicians to those areas is minimal. More practices are solo or partnership and lack the support found in group practices. Community health facilities and allied health professionals are rare, thereby restricting the diagnostic, therapeutic, and social service resources they can draw upon. (Rising, 1970)

Motivation

No mechanism exists to accurately assess and respond to the motivational needs of the differing types of general physicians. Lacking such a mechanism, the individual doctor must be stimulated to become active in his/her own behalf through being made aware of new knowledge and by being exposed to specially designed opportunities to fulfill their individual practice needs. (Brown, 1971; Pennington, 1979).

CME Program Content

CME offerings are seldom directly applicable to particular patient problems in general medical practice. This is due to a failure to link the topics of

---

* Requests for reprints should be sent to Muriel H. Bagshaw, M.D., American Institutes for Research, P.O. Box 1113, Palo Alto, California 94302.

1 This project was supported by contract N-01-HR-62932 from the Division of Lung Diseases, National Heart, Lung, and Blood Institute, N I H.

the course to the actual clinical problems of the practitioners, resulting in material which is either too esoteric and theoretical or too general to cover the range of particular clinical problems confronting the physician. (Wang, 1979; Mazamanian, 1977).

Evaluation

The hottest issue in CME is the lack of objective measures of the usefulness of the broad variety of programs offered. Problems in evaluation are associated with the long tradition within the profession to oppose outside interference in office practice. As more physicians are becoming knowledgeable about the advantages of providing well-documented and objectively auditable care, some improvement in this problem appears to be occurring. Yet the norm appears to be persisting: that the individual physician determines what should be done in any particular case, at least outside of institutional sites for medical care delivery. Evaluation, then, in the private setting requires an innovative approach and is a most sensitive issue. (Barro, 1973; Rubenstein, 1979; Bertram, 1977; Donabedian, 1978).

Methodology

The project to develop and evaluate a model CME program was divided into two phases; developmental and operational. In the first phase educational objectives were developed and materials produced and packaged. During the second phase the subjects were recruited, and the educational program field tested and evaluated.

Developmental Phase

Needs Assessment Procedures

To assess the practice needs of targeted physicians, three complementary approaches to needs assessment were utilized:

o Collecting Developmental Data. A field survey of the practices of a developmental sample of physicans in general practice or family practice in non-urban areas of California was conducted. Project personnel visited 30 practice sites in diverse underserved locations and interviewed each physician with regard to general CME issues, specific educational needs and preferences, and diagnostic and management problems in the area of asthma. These practices were recruited from the Stanford and Davis mid-level health practitioner Training Programs. A packet of office logs for recording all asthma patient contacts over an eight-week period were delivered at the time of the interview. These were to be mailed weekly to the project office. Eighty percent of these physicians completed the logs resulting in 406 office contact descriptions of asthma problems in rural practices.

o Obtaining Expert Input. Simultaneous with the field survey, a comprehensive literature search was conducted in consultation with specialists in pulmonary medicine. After collecting, reviewing, and cataloging current medical literature, a master protocol on the diagnosis and management of asthma was developed. A comprehensive topic outline on the diagnosis and management of asthma in adults was developed from input from the developmental samples case reports, current medical literature, medical specialist consultants, and the protocol.

o Rating of the Comprehensive Topic Outline. The topics in the outline were rated as to relative importance and relevance to clinical practice by a panel of 34 physicians in general practice and family practice, faculty in family medicine, and specialists in pulmonary medicine and allergy/immunology. Based on the results of the content ratings, the educational objectives for the CME program were identified.

These educational objectives were clustered into eight modules on the diagnosis and management of asthma in adults, with particular care given to developing a comprehensive, self-contained instructional presentation.

Preparation of Instructional Materials

The task of developing a comprehensive set of instructional modules was perhaps the most difficult and time-consuming aspect of the project. Five basic criteria for acceptable instructional presentations were:

- o The content was restricted to the identified educational objectives.
- o All information presented was to be documented by current acceptable medical publications.
- o The presentation was to be pertinent to the office practice of the generalist, not the specialist.
- o The information was to be clinically oriented rather than theoretical in nature.
- o The writing style was to be concise, using medical terminology familiar to the primary care physician.

The identified source materials, in most cases, could not be used in their original form for the following reasons:

- o The original sources were not intended for instructional purposes.
- o Sources were not often pertinent to the clinical practices of primary care physicians.
- o Sources varied greatly in organization and structure.
- o Sources varied greatly in style of presentation and use of medical terminology.

Medical writers were employed to develop texts that were directly related to the defined educational objectives and based on the reference materials selected by project staff. A bibliography was appended to each module to identify the references used in perparing the presentation. Careful editing was required to clarify and shorten the manuscripts.

Module packets included text, self-test, test annotation, and a critique form. Module titles were:

The Many Forms of Asthma, Making the Diagnosis of Asthma, Differential Diagnosis of Asthma, Assessing the Severity of the Acute Asthmatic Episode, Management of Asthma as a Chronic Disorder, Treatment of the Acute Asthmatic Episode, Involving the Patient in Self-Management, and Manual on Self-Audit.

Operational Phase

Recruitment of Participants

The method for enlistment of a representative group of practicing physicians to try out the model educational program entailed locating the target pool of individuals; random assignment of all individuals into one experimental and two control groups; and communication of motivational and enlistment materials to each of the groups.

Accurate location of the targeted pool was the most difficult of the three. First, the target criteria were chosen to include all private physicians in general or family practice in California who were located in towns under 20,000 located in medically underserved/rural Medical Service Study Areas (MSSA) in the state. Many problems were encountered in obtaining up-to-date lists, so that alternate methods were identified with the help of California State Health Department agencies. Details such as age, specialty (GP or FP), practice type (solo or group), and accurate addresses were essential. When these data were complete, they were recorded on individual file cards. A total of 694 individuals who met the criteria were found.

Next, the cards were cast into three groups in the ratio of 4:1:1, such that four stratification variables were equally assigned to all groups. Stratification was made for degree of medical underservedness (severe vs. moderate), date of M.D. degree (before 1967, after 1967), specialty (FP vs. GP), and region of the state (mountain/coast vs. valley/desert). Thus three pools were generated: Group E (Experimental = 466; Group C_2 (Control #2) = 114; and Group C_1 (Control #1) = 114.

Communication steps included accreditation for CME (Category I) from the California Medical Association, announcement of the project by notices in the state and county medical journals, and letters to Lung Associations and county medical society presidents. A brochure with a special project letterhead was professionally designed, detailing the plans and goals of the project. The brochure and a personal letter from the project director stressing the individualized nature of the program and opportunity for input from participants was mailed to Group E. Group C_1 and Group C_2 received letters asking for assistance in a research project on testing for knowledge of asthma and treatment.

In summary, all physicians in the state who were eligible (according to specific criteria) were identified and personally offered the opportunity to participate in the project. Balanced experimental and control groups were found by generating three independent target pools.

Operation of the Delivery System

Maintaining interaction with participants required a well-organized and executed delivery system. Confidentiality was assured by use of confidential code numbers. A master log was kept in which every contact was entered, and individual summary sheets were kept by code and updated regularly.

Of those who expressed interest in participation after the initial letter and one remailing, sufficient numbers were attained to field test the system. There were no differences in the four stratification variables among the three groups who dropped out and those who completed the project and control exercises. A total of 189 of 694 (28%) returned signed cards. In Group E, 72% (89 of 124) returned the pretest compared to Group C_2, 58% (21 of 36), resulting in an overall completion rate of 69%. Seventy physicians in Group E requested modules for study, representing 79% of those completing the pretest. Of these, 53 (67%) returned module tests and critique forms. The response to the posttest was 87% of those who studied modules in Group E and only 40% in Groups C_1 and C_2. Forty-three physicians in Group E completed all the modules as well as the posttest. The original goal for participation was for 50 in Group E and 15 in each of the control groups. By recruiting from a large pool, representation of the major stratification variables was maintained in the experimental groups.

Evaluation Results

Group E demonstrated statistically significant ($p = <.001$) gains in knowledge of asthma management while Groups C_1 and C_2 did not.

At the conclusion of the project, Group E physicians completed questionnaires on usefulness, format, and specific aspects of the system. Over 90% judged the system positively and would participate in similar programs again. More than 90% approved the specific content and style in each module.

It was demonstrated that a self-audit manual based on educational objectives was practical for use by busy primary care physicians.

Highlights of the lessons learned in the field testing of this system with a representative group of physicians in non-urban practice are:

o Needs assessment should cover specific patient problem needs as well as expert and potential user review of the topics proposed for the program.
o The learner groups should be relatively homogeneous for a program to be successfully tailored to their assessed needs.
o Resource materials must be carefully reviewed and abstracted if appropriate content is to be presented.
o Clearly specified educational objectives are essential for successful selection of material and for accurate evaluation of the program.
o The free-standing modular format improves motivation for the physician and forces clarity of presentation by the educators. High quality editorial expertise is critical to this task.
o Pretesting did not sensitize a control group to do better on a posttest than a posttest-only control group.
o Repeated personalized mail contacts between learners and delivery system project staff during the entire program was a successful motivational strategy.
o The problem of high drop-out rates in research designs among busy physicians can be solved by recruitment strategies that tap pools of potential participants and control for major variables in subjects and their locations.

References

Barro, A.R. Survey and evaluation approaches to physician performance measurement, Supplement, Journal of Medical Education, 1973, 48, 11, 1085-93.

Bertram, D.A. & Brooks-Bertram, P.A. The evaluation of continuing medical education: A literature review. Health Education Monographs, 1977, 5, 4, 330-362.

Brown, C.R. & Fleischer, D.S. The bi-cycle concept-relating continuing education directly to patient care. New England Journal of Medicine, 1971, 284, 88-97.

Donabedian, A. Needed Research in the Assessment and Monitoring of the Quality of Medical Care. (NCHSR Research Report Series, U.S. Public Health Service Publication No.78-3219) Hyattsville: National Center for Health Services Research, 1978.

Mazmanian, P.E. & Pennington, F.C. Needs assessment: Use of existing data sources, presented at AAMC RIME Symposium: Alternate Mechanisms to Assess Education Needs in Continuing Medical Education, Washington D.C. 1977.

Pennington, F.C. Continuing education, Letter to the editor, Journal of Medical Education, 1979, 54, 9, 752-3.

Rising, J.D. (Ed.) Proceedings of the Conference on Evaluation in Continuing Medical Education, University of Kanasas School of Medicine, August 1970.

Rubenstien, E. et al, Continuing medical education before audit, Medical Care, October 1979, 17, 1048-1053.

Wang, V.L. et al, Evaluation of continuing medical education for chronic obstructive pulmonary disease, Journal of Medical Education, 1979, 54, 10, 803-811.

Williamson, J.W. Improving medical practice and health care, Cambridge: Ballinger Publishing Co., 1977.

PATIENT CARE APPRAISAL IN THE AMBULATORY SETTING: EFFECTIVENESS AS A CONTINUING MEDICAL EDUCATION TOOL*

Wayne Putnam, M.D.** and Lynn Curry, Ph.D.
Dalhousie University, Faculty of Medicine

The overall objective of our project was to assess the impact of patient care appraisal (medical audit for education) with relevant interventions on physician performance in the ambulatory setting. A secondary objective was to test the value of participation in the process of selecting conditions to be audited and in the process of generation of criteria for those conditions by the physicians whose charts were to be reviewed.

For many years continuing medical education imitated undergraduate teaching with its primarily didactic classroom format for transferring information. Recent years have seen attempts to increase the relevance of continuing medical education (C.M.E.) for the practising physician by tying patient care evaluation into the educational planning and evaluation cycle, described as the Bi-Cycle concept by Brown and his colleagues. The evolution of this process into a quality assurance mechanism in hospitals has been hastened through adoption by accreditation agencies even though there has been little critical scientific work with regard to validity in improving patient care. A general assumption was made that there would be a natural extension of this process into care being given in the ambulatory setting and the investigators wished to test its usefulness before political pressure for accountability leads to the adoption of such a program in Canada.

BACKGROUND The conceptual foundations for Patient Care Appraisal (PCA) lie in the work of Miller, Williamson and their colleagues (1,2,3), developed by Brown and Fleisher (4) into a "Bi-Cycle" relationship between patient care and education. Davidson, Lein and Kelday (5) described how practising community physicians can participate in the selection of diseases to be audited and in the generation of criteria, the comparison of criteria with data of actual care and with the planning of educational or other corrective actions. It is this process applied to the ambulatory setting that forms the basis for the study described here.

Although there has been a general lack of critical evaluation, one exception is the Mandate Project (6), in which, for the 5 hospitals that completed the study, there was clear cut improvement in the quality of care being provided. We are not aware of any similar critical study of patient care appraisal in the ambulatory setting, and our project derives a great deal from the methodology and analytic approach of the Mandate Project.

Although a great deal of material has been published in the area of quality assurance, one of the best overviews was written by Brook, Williams and Avery (7). The authors concluded that attempting to develop an ideal standardized quality assurance system in the next few years is folly but

---

* This paper is based on research funded by Health and Welfare, Canada, NHRDP grant # 603-1075-44.

** Reprints: Dr. Wayne Putnam, Director, Division of Continuing Medical Education, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

there should be a series of experiments in different forms and that only the good should be kept. There are so many different health care delivery systems in North America that this statement is obviously true. Our attempts to develop and test a mechanism for tying C.M.E. to patient care in the ambulatory setting should provide one of the many answers needed to make progress in this field.

METHODOLOGY

A flow chart summarizing the study design has been included as Figure 1.

Sample The sample was selected from a list of all family physicians in Nova Scotia and New Brunswick who were in full time practice with access to adequate hospital beds for continuity of care and who had been in practice from five to twenty-five years. A letter was sent to all who met the entry criteria and sixteen were randomly selected from those who were willing to participate. They were randomly allocated to "treatment" and "control" groups. All project physicians were asked to keep a list of every condition seen in their office for two six-month periods and these diagnoses were coded by the International Classification of Health Problems in Primary Care (ICHPPC).

Selection of Conditions All of the diseases and conditions that were seen frequently enough to provide 50 episodes of care in a 6 month period were extracted from those lists, 50 having been chosen as the number with a strong likelihood of providing us with statistically significant results. From these, the diseases which met the six criteria for auditability laid down by Kessner (8) and the conditions described by Sibley (9) were chosen for potential involvement in the study. From this final list three were selected by the research team and two were selected by the "treatment" physicians.

Criteria Generation The optimal criteria of care for two of the above conditions were generated by the actively involved "treatment" physicians during visits to each by the principal investigator. These criteria were later collated by the research team and validated by a committee of peers and appropriate specialists called together for that purpose. Each physician was informed of the committee's deliberations and, if there were any variations from the physician's own set of criteria, he was provided with an explanation. The differences, which were all relatively minor, were accepted by the physicians involved.

The same committee of peers and specialists generated criteria for the three other conditions on the same occasion. Figure 2 shows the conditions involved in the study by source of selection and source of criteria. The fifth condition (Urinary Tract Infection) was included, without the awareness of the participants, in order to estimate the effect of participation in PCA on their practice generally.

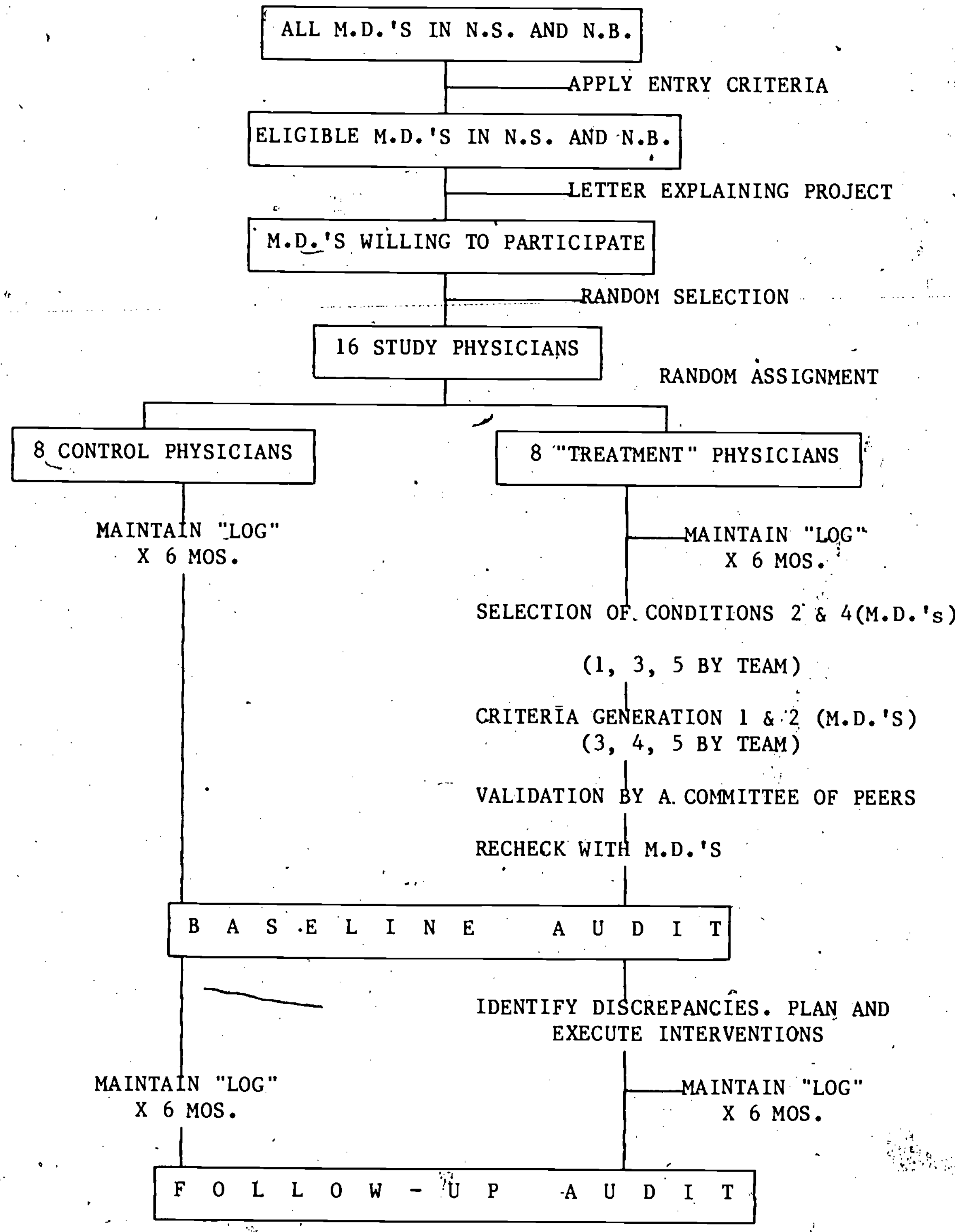


FIGURE 1.: STUDY DESIGN

		CRITERIA	
		SET BY PHYSICIANS	SET BY RESEARCH TEAM
CONDITION	SELECTION PHYSICIANS	BRONCHITIS	HEADACHE
	SELECTION RESEARCH TEAM	OTITIS MEDIA	HYPERTENSION URINARY TRACT INFECTION*

* Auditing for condition 5 (U.T.I.) not disclosed to participants until completion of study.

FIGURE 2.: STUDY DIAGNOSES

Baseline Audit A baseline audit was conducted during a visit to each practice by two health record analysts. Although they both had previous training and experience with record abstraction, a special program was carried out to standardize their skills at data abstraction from ambulatory records. (Inter-rater reliability test reached 97.6%).

Fifty episodes, or as close to 50 as could be achieved, of each condition were abstracted onto forms, with data coded to maintain confidentiality for both patient and physician. The data were then keypunched and stored in card form for later compilation and display.

Intervention Soon after the baseline audit, each physician was visited by the principal investigator and educational consultant and was shown his/her own results. These results were displayed as a percentage of episodes of care showing compliance with each criteria previously set, as in Figure 3. For example, in otitis media, usage of an appropriate antibiotic was a criterion of care and each physicians' own behaviour in that regard was displayed.

	YES		NO	NOT RECORDED		NOT APPLICABLE
Inspection of T.M.	98%	49		2%	1	
Antibioic Prescribed	100%	50				
Antibiotic Appropriate	100%	50				

FIGURE 3.: SAMPLE OF DATA DISPLAY FROM OTITIS MEDIA

During the discussion which followed, assistance was offered in relation to discrepancies between ideal and actual behaviour. Some requested literature to clarify an issue; for others, we provided educational material as an office aid to assist the physician in counselling patients. These individually-tailored packages were sent to each physician who had requested such assistance and after an appropriate period of time for reading the material, the physicians again began to keep daily lists of patients seen.

Follow-up Audit   A repeat audit was conducted six months later in the same manner as the baseline audit yielding data for comparison of behavior. At the conclusion of the two audits, it was obvious that there was a high degree of non-recording for many of the "optimal" criteria. This had been anticipated because we were told, at the interviews for criteria generation and reporting of results, that they did not record many of the activities they believed to be important. Consequently, we selected a smaller subset of essential criteria, defined by Sanazaro (10) as permitting "precise specification of a condition and known to be effective in producing the desired results". The analysis described was undertaken on the data derived from these essential criteria.

ANALYSIS   A conservative index of change was calculated for each condition for each physician:

$$\frac{(\text{\# of criteria improving by 15\%})\ (\text{audit 1 to audit 2}) - (\text{\# of criteria decreasing by 15\%})\ (\text{audit 1 to audit 2})}{\text{\# of criteria 15\% - 85\% at first audit}}$$

Computing this index allowed a reasonable representation of the true positive change in each physician. These index scores were analyzed for factor effects using multivariate analysis of variance.

RESULTS   Table 1 represents a compilation of the MANOVA results. The effect of being involved in the PCA process (A effect) appears robust. However, the effects of participating physicians choosing the audited conditions (D effect) or defining the audit criteria (E effect) appear non-existent in this data. Even focusing on the one condition that was both chosen by the physicians and defined by the physicians (G effect), there was no demonstrable difference compared to the other conditions. There was no difference in the concealed condition between treatment and control groups (F=.166).

	MEAN SQUARE	SUM OF SQUARES	DF	F	SIGNIFICANCE
A	2.06427	2.06427	1,42	13.20828	.001
	.15629	6.56403			
D	.01089	.01089	1,22	.04617	.832
	.23585	5.18876			
E	.05280	.05280	1,22	.22571	.639
	.23395	5.14685			
G	.07324	.07324	1,42	.35954	.552
	.20369	8.55506			

TABLE 1

DISCUSSION   If continuing medical education can be defined in its broadest sense as any activity which helps a physician improve care of patients, then PCA is an effective C.M.E. tool. Compliance with standards of care (both self generated and peer generated) increased when physicians were involved in the process. However, it is not at all clear what part of the process is primarily responsible for the change, whether it is the criteria setting, knowledge of deficiencies, discussions with principal investigator and educational consultant, or the packages.

Conversely, participation in the selection of the conditions or in the generation of criteria can not be shown with our data to have any effect on physician behaviour. We believe this is because the results are confounded by the substantial differences in the nature of the four conditions involved. Acute bronchitis and acute otitis media are both short term illnesses, relatively simple and straightforward as diagnostic entities and with regard to treatment. On the other hand, headache is a symptom of several different diagnostic etiologies and many different, yet acceptable, therapeutic approaches. Hypertension is a chronic condition characterized by the necessity for periodic screening procedures and therapeutic changes if indicated by changes in the natural progression of the condition.

CONCLUSIONS AND IMPLICATIONS C.M.E. directors should be aware of the potential of PCA as an alternative educational method for those physicians who enjoy planning their professional development with the assistance of critical self-assessments. However, more research is needed with regard to the value of involvement in criteria generation, knowing one's own practice deficiencies, and various types of educational packages. More investigation needs to be done to clarify which diseases and conditions seen in the ambulatory setting are appropriate for involvement in PCA.

REFERENCES

1. Williamson, J.W.; Alexander, M.; Miller, G.E. Priorities in Patient-Care Research and Continuing Medical Education. J.A.M.A. 204:303-308, 1968.
2. Williamson, J.W.; Alexander, M.; Miller, G.E. Continuing Education and Patient Care Research: Physician Response to Screening Test Results. J.A.M.A. 201(12):938-942, 1967.
3. Williamson, J.W. Evaluating Quality of Patient Care - A Strategy Relating Outcome and Process Assessment. J.A.M.A. 218(4):564-569, 1971.
4. Brown, C.R.; Fleisher, D.S. The Bi-Cycle Concept - Relating Continuing Medical Education Directly to Patient Care. N.E.J.M. 284 (Suppl):88-97, 1971.
5. Davidson, R.C.; Lein, J.N.; Kelday, J.M. Quality Assurance in the Provision of Hospital Care. Case Study. Audit Network Aids Continuing Medical Education. Hospitals 48:85-88, 1974.
6. Fleisher, D.S.; Brown, C.R.; Zeleznik, C.; Escovitz, G.H. Omdal, C. The Mandate Project: Institutionalizing a System of Patient Care Quality Assurance. Pediatrics 57(5):775-782, 1976.
7. Brook, R.H.; Williams, K.N.; Avery, A.D. Quality Assurance To-day and Tomorrow: Forecast for the Future. Ann. Int. Med. 85:809-817, 1976.
8. Kessner. D.M.; Kalk, C.E.; Singer, J. Assessing Health Quality - The Case for Tracers. N.E.J.M. 288(4):189-194, 1973.
9. Sibley, J.C. A Quantitative Approach to Quality of Patient Care. Paper presented at the Royal College of Physicians and Surgeons of Canada 45th Annual Meeting, Quebec City, 1976.
10. Sanazaro, P.J. Medical Audit, Continuing Medical Education and Quality Assurance. West. J. Med. 125:241-252, 1976.

PRECIS

CLINICAL PROBLEM SOLVING

MODERATOR: Geoff R. Norman, Ph.D.
McMaster University, Department of Family Medicine

EXAMINATION OF THE EFFECTS OF STRUCTURED SMALL GROUP FORMATS ON MEDICAL STUDENTS' PROBLEM-SOLVING PERFORMANCE

This paper examines the use of census and force field analysis in instructor-led seminars. The results indicate no differences between student groups who experienced census and force field analysis and students groups who experienced more traditional seminar formats with respect to their self-confidence in problem solving ability and performance on PMPs.

CLINICALLY RELEVANT PROBLEM SOLVING EVALUATION IN PRECLINICAL MEDICAL EDUCATION: A STUDY OF ALTERNATIVE APPROACHES

Undergraduate medical students completed objective tests and problem solving exercises for the clinical problems of fever and chest pain. Although results show no relationship between performance on problem solving and objective questions, they do indicate that the problem solving exercises are appropriate instruments for assessing application of the clinical inquiry model.

PROBLEM SOLVING ANALYSIS: A PIAGETIAN STUDY

A Piagetian Study of the problem solving abilities of fifty-nine sophomore medical students using twelve Piagetian mediated/written tasks, comprising four logical schemata (i.e., combinatorial, propositional, probabilistic and proportional reasoning). All subjects were found to be transitional formal or fully formal on these tasks.

Examination of the Effects of Structured Small Group Formats on Medical Students' Problem-Solving Performance*

James M. Shumway, M.Ed.
James F. Donohue, M.D.

School of Medicine
University of North Carolina at Chapel Hill

Purpose

Instruction during the clinical phase of medical school is of utmost importance to medical students. In addition to being critical to an essential aspect of their education, it has significant potential for research and subsequent instructional improvement. It is probable that medical students learn diagnostic problem-solving skills by modeling the behavior and approaches used by the clinical instructors they encounter (1). Clinical instructors probably try to model what they consider to be appropriate problem-solving approaches applicable to each patient's problem. Medical students are expected to adopt and apply the skills they see clinical instructors use, with little or no formal training in problem-solving. There has been a practical concern noted among medical educators for finding ways to improve teaching and to measure the important physician characteristic of skill in medical problem solving (2).

Discussion of the process of medical problem solving by Elstein and his colleagues (3) suggests that divergent thinking is needed for effective hypothesis generation, and analytical and evaluative thinking is needed for effective hypothesis and data evaluation. The problem investigated in this study was to determine whether structured small group instructional formats in instructor-led seminars enhance advanced undergraduate medical students' diagnostic problem-solving skills. Medical students who have had exposure to census and force field analysis might be expected to develop a more clearly defined process of medical problem solving. The two questions central to this study were:

1. Do medical students who participate in instructor-led seminars that employ census and force field analysis demonstrate problem-solving performance, as measured by responses to written patient management problems, that is different from that of medical students who participate in traditional instructor-led seminars?

2. Do medical students who participate in instructor-led seminars that employ census and force field analysis demonstrate a level of confidence in their skill to solve patient problems, as assessed by their expressed self-confidence in their problem-solving ability, that is different from that of medical students who participate in traditional instructor-led seminars?

Literature Review

Diagnosis is an essential component of medical problem solving. Much of medical problem solving is focused on the differential diagnoses that students

---

Address correspondence and requests for reprints to first author at the Division of Educational Development, University of Kentucky, Lexington, Ky., 40536.

*This project was supported in part by a Public Health Service Pulmonary Academic Award (#K07HL00449) from the National Heart, Lung and Blood Institute, Dr. Donohue.

reach and apply as explanations of patients' problems. Weed (4) emphasized an approach which has the student or physician acquire a comprehensive data base. The Weed approach has been included in "Physical Diagnosis" courses. It teaches students to accurately record patient data on his/her medical record (5-6). Morgan and Engel (7) suggested that incoming data about a patient serves as a guide or selection process to subsequent data collection. Their suggestion is supported by the results of the Medical Inquiry Project (3) which concluded that initial diagnostic hypotheses tend to guide the subsequent data collection process, before a final diagnosis is reached.

Theory holds that hypothesis generation is a significant component of the diagnostic process and that initial diagnoses are usually generated early and from limited data (8). Research undertaken to describe the general process that physicians employ to generate initial problem formulations, hypotheses, or tentative diagnoses suggests the use of divergent thinking strategies in the initial problem formulations made by physicians. Allal (9), in her descriptive analysis of how practicing physicians reached initial problem formulations of patient's problems, sought to determine if they followed a convergent strategy or either of two divergent strategies. The convergent strategy meant that they attempted to come up with an initial problem formulation that accounted for all of the data. The first divergent strategy was that physicians attempted to think of as many formulations as possible that fit the observed data. The second and more modest divergent strategy was whether they attempted, with each formulation generated, to think of other possible formulations. Upon her examination of convergent versus divergent strategies, the physicians revealed that they used convergent strategies less frequently than divergent strategies. She speculated that the reason physicians rarely used convergent strategies of problem formulation during the initial four to six minutes of a workup was because it could entail the risk of premature closure (9). It appears that divergent strategies could reduce the error of premature closure. This could account for the preference of divergent strategies by experienced physicians.

Evaluation of hypotheses that physicians derive probably can be explained by simple additive models of decision making (10). Simple additive models of decision making involve a listing of factors for and against a decision, often with the assignment of differential weights to each factor according to their perceived importance, and then summing the values of each column. An example of differential weighting of the factors for or against a particular treatment, for example, might include weighting of those factors based on the cost of the treatment, whether it is potentially harmful to the patient, and whether the results of the treatment will cause immediate improvement in the patient's condition. Elstein, et. al. (3) tested three linear models through the diagnostic judgment of physicians. They found that basing decisions on negative weights alone was the poorest predictor. Basing decisions on positive weights alone did an average-to-good job in predicting, depending on the situation at hand. Performance using a model which based the decision on the maximum difference of positive cues minus negative cues was virtually indistinguishable from the model that based decisions on positive weights alone. They also found that listing factors for, listing factors against, or listing those factors that did not help did at least as good a job of describing decision making behavior as did more complex differential weighting systems. Based on those results, they concluded that a simple additive linear model of decision making may be helpful to physicians in making more accurate diagnostic decisions.

The process of teaching medical problem solving to medical students involves the designing of instructional formats that enhance and foster the learning experience as well as the skills in divergent and analytical thinking. In a review of the literature of studies on teaching that have measured general problem-solving ability, the results favor discussion over lecture (11). Research indicates that small group learning is more effective than lecturing for producing higher-level productive thinking (12). It has been suggested that small learning groups can be structured via normative instructions to produce certain kinds of problem solving (13). Jaffe (14) found that small learning groups of nursing students who received normative instructions in nominal group process and in consensual decision making evidenced more productive thinking that did student groups who responded only to open-ended questions in a discussion. She concluded that one can structure the group process to facilitate productive thinking. Since the teaching of medical problem solving to medical students would most likely require an intensive learning experience, instructor-led seminars using structured small group formats may be an appropriate instructional format to enable medical students to develop problem-solving skills.

Method

This study was conducted as a field experiment. The population was the fourth year medical student class at the University of North Carolina at Chapel Hill School of Medicine (UNC-CH). Thirty-six medical student volunteers comprised the sample of subjects. These students were randomly assigned to six seminars. Each seminar met for approximately one and one-half hours weekly for three weeks. A different patient case in pulmonary medicine was considered each week. The treatment condition of census and force field analysis, small group instructional strategies, were randomly assigned to three of the seminars.

Census is a small group instructional format, similar to brainstorming techniques which you may be familiar, which facilitates the initial stage of problem solving that involves the listing or identification of existing solutions, issues, logical alternatives, etc. (15). Unlike brainstorming, however, it emphasizes the use of logic and realism from the very beginning of the process whereas brainstorming includes the realm of fantasy. It promotes divergent thinking which involves the identification of information (not given in the problem statement) using the data provided as a basis (15). Divergent thinking facilitates the identification of logical alternatives for the solution of a problem. The data generated usually come from some pre-existing set of information, (i.e., that data which are collected in the history, physical examination and laboratory studies). The census procedure, as carried out in a small group instructional setting encourages individuals to withhold evaluation of other members' hypotheses or logical suppositions during the generating phase; encourages individuals to utilize the information offered by other members; and stimulates the identification of additional logical possibilities as individuals call out their speculations in spontaneous fashion. In this respect, census is used as an instructional technique in instructor-led seminars to enhance opportunities for learning the skills of initial hypothesis generation, given limited data about a patient.

Force field analysis is a structured small group format that can be used to facilitate medical students' hypothesis evaluation performance by using a simple additive model for decision making by enhancing analytical thinking.

The purpose of force field analysis is to help individuals move from divergent thinking to analytical thinking by applying logic and reason to a problem, and to analyze a problem in terms of the factors for and against the adoption of a particular solution. It is used in instructional settings to elicit analytical thinking in individuals with regard to a possible solution to a problem. Analytical thinking "emphasizes the breakdown of material into its component parts and the determination of the relationships among the parts and the way in which they are organized and relate to the whole (15)." It is a systematic approach for examining a proposed solution to a problem by examining the forces for and against the adoption of the proposed solution and for finding ways to either increase the forces for or to reduce the forces against a particular solution. Often force field analyses of several competing proposed solutions can be competing proposed solutions can be compared to determine the best one. Individuals participating in a force field analysis are encouraged to examine both sides of an issue in determining its' characteristics.

The control condition, traditional instruction, was randomly assigned to the three remaining seminars. In the control condition seminars, the instructor asked students to read the case. One student usually read the case out loud as the others followed the text and read along silently. After the case was read, the instructor asked open-ended questions about it. Such questions, for example, were: "On the basis of the information obtained in the history and physical, classify this attack (Asthma) as mild, moderate, or severe. List the evidence supporting this assessment."; "You decide to treat the patient before ordering time consuming additional laboratory tests. Outline your initial therapy and briefly indicate the mechanism of action of the therapy chosen."; "List the possible mechanism for bronchospasm in this patient. Rank them as the most probable to least probable."; and, "Why do asthmatics seem to have more difficulty at night?"

The criterion instruments used to measure students' medical problem-solving were: (a) a 24-item summated rating scale that measured medical student's expressed self-confidence in their medical problem-solving ability, and (b) four written patient management problems (PMPs) in pulmonary medicine. The expressed self-confidence instrument was adapted form a scale developed by Berger (16). The reliability of the instrument assessed by Cronbach's coefficient alpha (17) was .86 with the medical student sample. The 4 PMPs in pulmonary medicine were chosen from currently published sources of written patient management simulations (18-20). The problems were evaluated independently by five faculty members and were judged to be representative of common problems involving pulmonary diseases. The scoring and appropriateness of each item for each PMP was determined by the local faculty according to the local practice of clinical medicine. All PMPs followed a linear format and contained sections on data collection, patient management and, in one problem, a section on pathology. The reliability (21) of the four PMPs using Cronbach's coefficient alpha (17) was .76, when individual items were the unit of analysis, for the sample of 36 medical students. The validity of the PMPs were determined by correlating scores of similar sections across problems and comparing them with correlations of scores of different sections. A stronger relationship was found between similar sections than between different sections. Five scores (proficiency, errors of omission, errors of commission, efficiency and competence) were computed for each subject on each section of each of the four PMPs (22). The subject's overall scores for data gathering and management were calculated as the sum of those sections on all four PMPs.

The design used for statistical analysis was hierarchal, where one or more of the treatments is nested instead of being crossed as in classification or cross-treatment experiments. The use of a hierarchal design takes into account that seminars, per se, may contribute significantly to the total variation among the subject's scores (23). In this respect, seminars conceptually were the unit of analysis. An F ratio was the test statistic used to determine if a significant difference (alpha = .05) existed between the treatment and control groups.

Results and Discussion

The analysis failed to show any statistically significant difference between the treatment and control groups with respect to subjects' expressed self-confidence in their medical problem-solving ability. The fact that no differences were found in self-confidence was disappointing. It was hoped that medical students' attitudes in the treatment seminars which encouraged divergent and analytical thinking in the solution of patients' problems would strengthen individual's self-confidence. However, students' self-confidence with respect to their medical problem-solving ability may already be fairly well established by the time they reach the clinical phase of their training. Consequently one might speculate that the lack of a difference between the two groups might be because the treatment condition was not strong enough or long enough to change the pattern of evaluation and feedback they normally get from their peers and instructors in the seminar setting.

The analyses of the data gathering, patient management and combined sections for all 4 PMPs failed to detect any statistically significant differences between the treatment groups and the control groups for any of the 5 scores. This result was unexpected since the literature reviewed on instructional approaches found structured small groups more effective than traditional group instruction for fostering problem-solving skills (11-14). Reasons that might have attributed to the lack of difference between the groups in overall problem-solving performance are: (a) the treatment seminar on problem solving in pulmonary medicine may have appeared too late in the curriculum for medical students, who, by the middle of their senior year, may have already formulated their unique individual medical problem-solving strategies; (b) the duration of the training and length of time spent in the treatment seminars may not have been long enough to alter the unique problem-solving abilities of individual students; (c) there may, in fact, be no difference in medical students' problem-solving strategies and the divergent and analytical strategies for problem solving suggested in the census and force field analysis seminars; and, (d) the PMPs may not have been a sensitive enough instrument to illuminate differences produced by the educational treatment.

The implications of this study suggest that further research be conducted on the use of instructor-led seminars for teaching medical problem solving. Other research methodologies (e.g., ethnographic approaches and content analytic procedures) should be considered to help us understand how medical students develop problem-solving skills.

References

1. Miller, et al. Teaching and Learning in Medical School. Cambridge, MA: Harvard U Press, 1961. Chapter 13: Techniques for Measuring Knowledge. Chapter 8: Basic Techniques of Instruction. HSL W18 M648 1961.
2. Jasm, H.A. "A Study of Medical Teaching Practices." JMC, 37:1258-1284, 1962.
3. Elstein, A.S., Shulman, L.S. and Sprafka, S.A. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, MA: Harvard U Press, 1978.
4. Weed, L.L. Medical Records, Medical Education and Patient Care. Cleveland, OH: Case Western University Press, 1971.
5. Weed, L.L. Medical records that guide and teach. New England Journal of Medicine, 278, 593-599, 652-657, 1968.
6. Weed, L.L. Medical Rewards, Medical Education and Patient Care. Cleveland, OH: The Press of Case Western Reserve University, 1969.
7. Morgan, W.L. and Engel, G.L. The Clinical Approach to the Patient. Philadelphia, PA: W. B. Saunders Co., 1969.
8. Sprafka, S.A. The effect of hypothesis generation and utilization on certain aspects of medical problem solving. Unpublished Ph.D. dissertation, Michigan State University, 1973.
9. Allal, L.K. Training of medical students in a problem-solving skill: the generation of diognostic problem formulations. Unpublished Ph.D. dissertation, Michigan State University, 1973.
10. Dawes, R.M. and Corrigan, B. Linear models in decision making. Psychological Bulletin, 81, 95-106, 1974.
11. McKeachie, W.J. and Kulik, J.A. Effective College Teaching in F.N. Kerlinger (ed). Review of Research in Education. #2, Itasca, IL: Peacock, 1975.
12. Lloyd, K.E., Garrington, W.K., Lowry, D., Burgess, H., Euler, H.A. and Knowlton, W.R. A Note on Some Reinforcing Properties of University Lectures. J. of Applied Behavioral Analysis, 1972, 5, 151-155.
13. Hall, J. and Williams, M.S. Group dynamics training and improved decision making. J. of Applied Behavioral Science, 1970, 6, 39-68.
14. Jaffe, Joyce. The Relationship of Structured, Student-Centered Small Group Instruction Formats to Higher-Level Thinking in Collegiate Learning Groups. (UNC-CH, unpublished dissertation 1978).
15. Jaffe, J. and Watson, E.R. Resource Packet for Small Group Instructional Skills and Strategies. Unpublished manuscript, Office of Medical Studies, University of North Carolina, 1976.
16. Berger, E. The relation between expressed acceptance of self and expressed acceptance of others. J. of Abnormal Social Psych, 1952, 47, 778-782.
17. Cronbach, L.J. Coefficient Alpha and Internal Structure of Texts. Psychometrika, 1951, 16, 297-334.
18. Webster, G.D. and Cassileth, P.A. Guide to Patient Management Problems in Medical Examinations. New York: Elsevier North-Holland, Inc., 1978.
19. Hubbard, J.P. Measuring Medical Education. Philadelphia, PA: Lea and Febiger, 1978.
20. McGuire, C.H., Soloman, L.M. and Forman, P.M. Clinical Simulations. 2nd ed. New York: Appleton-Century Crafts, 1976.
21. Lewy, A. and McGuire, C. A Study of Alternative Approaches in Estimating the Reliability of Unconventional Tests. A paper read at the Annual Meeting of the AERA, Feb. 18, 1966, Chicago.
22. McGuire, C.H., Solomon, L.M. and Bashook, P.G. Construction and Use of Written Simulations. New York: The Psychological Corp., 1976.
24. Kirk, R.E. Experimental Design: Procedures for the Behavioral Sciences. Belmont, CA: Brooks Cole Publishing Company, 1968.

CLINICALLY RELEVANT PROBLEM SOLVING

EVALUATION IN PRECLINICAL MEDICAL EDUCATION

A STUDY OF ALTERNATIVE APPROACHES

Sarah A. Sprafka, Ph.D.
Verda M. Scheifley, Ph.D.

Office of Medical Education, Research and Development
Michigan State University[1]

In a continuing effort to achieve desirable levels of both validity and reliability, clinically relevant performance evaluation in medical education has walked a line between objective tests--such as multiple choice tests--which are usually reliable but not always seen as valid, and more or less extensive case work-ups which can be high on content and face validity, but may suffer as evaluation instruments from lack of reliability. The unreliability of these tests thwarts any effort to assess their criterion related validity.

One source of unreliability in the case work-up approach to evaluation is the phenomenon of case specificity or inconsistency in performance across cases. This source of inconsistency was discussed by Elstein et al. (1978) and is currently being further investigated (Berner and Engel, 1979).

Background and Theoretical Framework

Until now, achievement of the objectives for the problem-oriented components of the preclinical curriculum at Michigan State University's College of Human Medicine has been assessed in two major ways. Students have taken objective examinations consisting of questions focused on the learning objectives and they have been asked to complete problem solving exercises which are extensive case work-ups requiring them to demonstrate ability to use the problem solving model outlined by Elstein et al. (1978), to use the Problem Oriented Medical Record, and to demonstrate knowledge of clinical, basic and behavioral science information relevant to the case. The theoretical basis and structure of these exercises is described in detail elsewhere (Sprafka et al., 1979). Analysis of performance on these cases has suggested that there may be a lack of consistency from problem to problem for any given student or group of students. For example, results of data sets collected over the 1978-79 academic year indicated low or no relationship between ratings on various problem solving exercises. Correlation coefficients ranging from .09 to .33 were obtained on scores for one group of students (N = 70). For another group (N = 16) the result was r = .34. Furthermore, analyses have shown that generally, there is no relationship between performance on problem-solving cases and objective examinations. For example, correlations between problem solving cases and objective exams for the groups mentioned above were from -.02 to .28 and from -.12 to .22 respectively.

---

[1] Requests for reprint should be addressed to Sarah Sprafka, Office of Medical Education, Research and Development, A-214 E. Fee Hall, Michigan State University, East Lansing, Michigan 48824.

The lack of relationships is seen to be due in part to certain psychometric properties of the exercises and in part to the nature of the task. The problem exercises, though generally addressing the learning objectives as described above, may have placed emphasis on different aspects of the problem solving process or on different objectives depending on the problem presented. Thus, as tests, they could not be readily compared to each other. Furthermore, students' responses were rated on a scale from 1-4 which resulted in actual ratings for most students of 3 or 4. The low variance in the ratings contributed to reducing the degree of correlation between the variables. This difficulty was compounded by the fact that there was a lack of consistency in the ratings, as various clinical preceptors evaluated different groups of students. Inconsistency in performance across problems may also be attributable to the nature of the problem solving task. It is possible that regardless of the psychometric properties of the instruments, knowledge factors or the structure of the problem solving situation may be the basis for problem specific differences.

Maatsch et al. (1978) achieved high reliability on clinical simulations with careful design and rater training, and performance on objective tests correlated highly with performance on clinical problem solving exercises. Thus it appears that instruments can be developed which are reliable, and positive relationships between simulations and clinically relevant objective tests can be attained. It still remains for us to gain a better understanding of case specific differences.

The study outlined here proposes to address questions relating to the relationship between problem solving examinations and objective tests, and case specificity as they occur at the preclinical level. More specifically:

1. What is the relationship between performance on the problem solving portions of a case work-up and portions of the exercise which assess content knowledge specific to the case?
2. What is the relationship between performance on problem solving and case specific content questions and performance on an objective test geared to a broader set of learning objectives of which the case is a sample?
3. Assuming that a general hypothetico-deductive approach to problem solving is applied in all cases, what, if any specific components of this process can be identified as salient in any given case as expressed in a set of carefully constructed comparable paper simulations?
4. How are the components of the problem solving process applied differently in different problems?

Method

Problem solving exercises were developed for two medical problems: fever and chest pain. Each case consists of two identifiable components: 1) essay questions emphasizing the Problem Oriented Medical Record format wherein students demonstrate their ability to use clinical data to solve and manage the problem and 2) objective questions relevant to the case. Each of these components is equally represented in each case, making the cases comparable with regard to format and variable only in the problem presented.

Case development was done by third and fourth year medical students. At the same time as the cases were developed, a criterion, or set of acceptable answers, was also created. All of the case and criterion materials were extensively reviewed and revised as necessary by clinical faculty members to insure accuracy of clinical content. Once the case and criteria were completed, a rating form was designed for use as a mechanism for recording judgments about student performance on the essay portion of the case. Responses were rated on an eight-point scale, where points along the scale range from clearly unacceptable (1) to superior (8).

During the fall term 1979, second year medical students devoted three weeks to studying materials associated with each problem - fever and chest pain. For each problem they were given the problem solving exercise as an in-class examination and were allowed four hours to complete it. In addition, students took an objective (multiple-choice) examination covering the learning objectives associated with the two problems (KR-20 = .77).

Evaluation of students' responses to the essay portion of each case was done by an advanced medical student familiar with the focal problem materials, the case, and the criteria and who was trained extensively by a clinician (r=.8). Objective items on the case and on the multiple choice exam were scored right or wrong. The percent of questions correct is the measure of student performance. Since students may progress through the curriculum at their own pace, more had completed the fever problem (N=42) than the chest pain problem (N=27) by the time the data were analyzed.

The relationship between problem solving ratings and objective test questions (questions 1 and 2) is examined by correlation and regression analysis. Where the correlation coefficient between scores on the objective items and the problem solving case is significantly different from zero, estimates of prediction equations between the two formats are found to determine those aspects of one format which best predict performance on the other.

Factor analysis is used to analyze the problem solving components (questions 3 and 4). The extent to which the hypothetico-deductive model of problem solving is used by students in solving the problems is determined by inspecting the structure of the correlation matrix for item ratings within each case. Factor loadings give an estimate of the relative impact of each item on the components of the problem solving model for each case. Differential application of the problem solving model across cases is examined by comparing the factor structure of the correlation matrices for each one.[2]

Results

Measures of correlation between performance on the essay portion of the case work-up, objective items on the case, and the multiple choice exam are reported in Table 1. There is no relationship between scores on objective case items with ratings on essay items or scores on the multiple choice exam for either case. For the chest pain case, there is a low positive correlation between performance on essay items and the multiple choice exam, although a similar relationship was not significant for the fever case. The regression equation pre-

2 The computer program used to analyze these data is the Statistical Package for Social Scientists (SPSS).

dicting performance on the objective test from the ratings on chest pain ($\hat{\beta} = 4.25$, std. error = 2.17) was marginally significant ($p < .06$) with $R^2 = .13$. This indicates a weak prediction model between the two variables.

TABLE 1

Correlation Coefficients
Multiple Choice Exams and Problem Solving Cases

	FEVER			CHEST PAIN		
	A	B	C	A	B	C
A. Multiple Choice Exam	1.00			1.00		
B. Case Objective Items	.14	1.00		-.12	1.00	
C. Case Essay Items	-.01	.06	1.00	.37*	.19	1.00

* $p < .03$

To assess overall similarity in performance across cases, mean scores were correlated. The relationship between ratings on the fever and chest pain cases was not significant (r=.24, N=27).

Factor analyses were performed on the essay items of each case to determine the components of the inquiry process represented in them, and to ascertain the nature of case specific differences. For the fever case, four factors had an eigenvalue of 1.0 or greater and together accounted for 61.2% of the variance. For the chest pain case there emerged five factors with an eigenvalue of 1.0 or greater accounting for 72.4% of the variance. Abbreviated statements of rating items and their respective factor loadings for each problem are shown in Table 2.

In the analysis for the fever case, a factor relating to assessments and plans (cue interpretation) emerged clearly. Other factors that emerged for the fever case related to development of a cue list for a progress note (Factor 2); problem formulation and cue interpretation (Factor 3); and planning the focus of the history (Factor 4). The results of the factor analysis for the chest pain problem are more difficult to describe. Factor 1 shows moderate to heavy loadings on items relevant to cue interpretation (explaining the cause of pain using available data and diagnostic and management plans). This is similar to Factor 1 for the fever problem though not as clear cut. Other factors that emerged for the chest pain problem are relevant to problem formulation and hypothesis generation (Factor 2); prediction of findings for possible causes of chest pain (Factor 3); prognosis (Factor 4); and a fifth factor that seems to combine data interpretation and emergency management measures.

Conclusions and Implications

With regard to questions 1 and 2 above, we must conclude that, similar to past studies, we found little or no relationship between problem solving items and objective items. This despite careful design, which resulted in increased score variance, and increased reliability.

In regard to questions 3 and 4, the results of the factor analysis suggest that components of the inquiry model can be identified in these more carefully

TABLE 2

FEVER

Factor Analysis
N=42

Item	Factor 1	Factor 2	Factor 3	Factor 4
	Factor Weightings - Varimax Rotation			
Relate findings to dx criteria	.760*	.098	-.384	-.508
Plan history	.314	.045	-.403	.644*
Relate data to possible causes	.410	.329	.484*	-.029
Problem formulation	.101	-.118	.451*	-.063
Assessment (ddx and cue interp.)	.583*	-.043	.285	-.035
Diagnostic plan	.543*	.006	.132	-.037
Progress Note: Retitle problem	.019	.274	.017	-.116
S and O data	-.052	.925*	-.186	.088
Update assessment	.753*	-.028	.144	.107
Dx plans	.470*	.050	-.086	.237
Pt. Ed. Plans	.014	.047	-.002	-.215
Eigenvalue	2.76	1.54	1.34	1.10
Cumulative % of variance	25.0	39.1	51.2	61.2

CHEST PAIN
N=27

Item	1	2	3	4	5
	Factor Weightings - Varimax Rotation				
Interpret hx for ddx	-.126	-.039	.295	.063	.426*
Emergency mgmt. decisions	.295	-.010	.056	.207	.845*
Relate data to possible causes	-.059	-.043	.028	-.220	.412*
Predict findings for possible causes	.114	.088	.811*	-.015	.137
Update MPL	-.136	.977*	.152	.019	.084
Assessment (ddx and cue interp.)	-.229	-.175	.254	.535*	.107
Dx and mgmt. plans	.628*	-.180	-.010	.347	.195
Patient education plans	.246	-.301	.481*	.070	.070
Explain cause of pain	.860*	.056	.311	-.238	-.135
Prognosis	.137	.026	-.080	.788*	-.085
Explain prognosis to pt.	-.043	-.307	.158	.068	.111
Eigenvalue	2.33	1.61	1.56	1.30	1.17
Cumulative % of variance	21.2	35.8	49.9	61.7	72.4

designed, more reliable exercises. The components that emerge most clearly are those relating to cue interpretation, including assessment of problems using available data, and planning for data gathering focused on hypotheses. The hypothesis generation component of the inquiry model is also identifiable as a performance factor. The data gathering and hypothesis evaluation components were not easily identifiable.

The low correlation between mean performance on the two problems suggests problem specific differences. On the surface the chest pain case appeared less diffuse and vague than the fever case. For the chest pain case the student's major task was to differentiate between two diagnoses (myocardial infarction and angina pectoris), and to manage the patient in the acute phase of his illness. The fever problem presented as a vague fever of unknown origin in an otherwise healthy young man. The student's major task was to try to diagnose the patient's illness. Given this, one would expect the diagnostic and management portions of the chest pain problem to cluster together. Since the fever problem focused on diagnosis, one would expect components of the inquiry model to emerge, but perhaps in a less orderly fashion than anticipated for the chest pain problem. The factor analysis suggests the opposite may be true. For the fever case, components of the inquiry model cluster together nicely, with the cue interpretation element being the most identifiable. However, for the chest pain case, although aspects of management emerged reasonably clearly (e.g. prognosis and emergency management), the components of the inquiry model did not cluster together, but were distributed throughout the factor structure.

This study has implications for the evaluation of clinical problem solving at the undergraduate level, as well as for future research. The results are encouraging for evaluation, suggesting that if the evaluation instruments currently in use continue to be carefully designed and reliable, they can identify varying levels of skill in applying the clinical inquiry model by undergraduate medical students. The results suggest that more study is needed to better understand why there appears to be no relationship between performance on problem solving exercises and objective tests, and to ascertain the basis for differences between problems, especially when they are counterintuitive.

REFERENCES

Berner, Eta S., Assessment and evaluation of medical problem solving. Paper given at the Conference on the Role of Problem Solving in Medicine, Smuggler's Notch, Vermont, October 18-21, 1979

Elstein, A.S., Shulman, L.S., and Sprafka, S.A., Medical Problem Solving: An Analysis of Clinical Reasoning, Cambridge, Harvard University Press, 1978

Maatsch, J.L., Downing, S., Sprafka, S.A., and Holmes, T., Toward a testable theory of physician competence: an experimental analysis of a criterion referenced specialty certification test library. Symposium given at Conference on Research in Medical Education, Washington, D.C., October 28, 1978

Sprafka, S.A., Elstein, A.S., and Parmeter, J.T., Patient centered problem solving evaluation in undergraduate medical education: Development, management, and theoretical perspectives, in Medical Education Since 1960: Marching to a Different Drummer, A.D. Hunt and L.E. Weeks, Eds., The Michigan State University Foundation and The W.K. Kellogg Foundation, East Lansing, 1979

PROBLEM SOLVING ANALYSIS
A PIAGETIAN STUDY*

JAMES P. HALE, PH.D.
UNIVERSITY OF SOUTH DAKOTA SCHOOL OF MEDICINE

INTRODUCTION

The theory of intellectual development proposed by Jean Piaget is stage dependent (Inhelder and Piaget, 1958). According to his theory, as a person develops and matures he passes through four stages of intellectual development, the first three of which are tied directly to the concrete world in which the individual is existing, and the fourth, the culminating stage of Piagetian theory, the individual is able to go beyond these concrete interactions with his environment and utilize a new formal hypothetic-deductive reasoning capability. This stage of intellectual development is referred to as the formal level of intellectual development and is characterized by the employment of a set of operational structures based on propositional logic. Inhelder and Piaget in describing formal thought added that besides verbal reasoning (propositional logic), formal thought "...also entails a series of operational schemata which appear along with it" (1958, p. xxii). These formal logical schemata (concepts which manifest themselves when the formal subject is faced with particular kinds of data, but do not ordinarily manifest themselves otherwise) include combinatorial operations, proportional reasoning, probabilistic reasoning, the scheme of mechanical equilibrium (equality between action and reaction), and four others. According to Inhelder and Piaget (1958), propositional logic and these formal logical schemata develop concurrently during adolescence (i.e., beginning at about age eleven).

The ability to use propositional logic and the acquisition of formal logical schemata is of considerable importance to the medical students. The understanding of proportional relationships, for example, is essential to the medical students when confronted with the numerous biological and physical concepts and principals such as diffusion, CO_2 concentrating and respiration rate, the low of definite composition, etc. Combinatorial logic is required for the understanding required in solving human genetics and biochemical mechanism problems. Probabilistic reasoning is required when applying epidemiological data to the differential diagnosis of smallpox or lung cancer. All in all, the ability to use propositional logic and the formal logical schemata not only play an important role in the medical student's ability to understand the scientific content of medicine, but also plays an important role in the process of medical problem solving (i.e., differential diagnosis).

PURPOSE

The purpose of this study was to investigate the development of propositional logic and three formal logical schemata in adults enrolled in their second year of medical school. Twelve formal-level Piaget-related tasks were identified and/or developed for use in this study. In addition to assessing the use of propositional logic, the study was designed to assess the level of intellectual

*Request for reprints should be directed to James P. Hale, Co-director of Office of Educational Evaluation, Research and Service, School of Medicine, The University of South Dakota, Vermillion, South Dakota 57069.

development of fifty-nine (of 65 possible volunteers) second year medical students within three of Piaget's eight formal logical schemata: Combinatorial Operations, Propositional Reasoning, Proportions, and Probability.

According to Inhelder and Piaget's (1958) framework, the onset of formal operations relies on the subject's ability to view the world as a whole, that is, not only concretely and abstractly, but hypothetico-deductively when confronted with a complex problem. If a subject is able to think in terms of all possible combinations when confronted with a problem, he is exhibiting a combinatorial ability. If, also, the subject is able to identify and control variables within a hypothetical experimental setting, the subject is exhibiting a principal rule of propositional logic. The combinatorial ability along with the ability to apply the common rules of propositional logic often manifests itself in the formal thinker when confronted with complex problems to solve. The logical schemata of probabilistic and proportional reasoning also require the use of combinatorial and propositional logic. They differ only in the way this logic is applied to the solution of problems. Specifically probabilistic reasoning requires that the "subject's deduction begins with possibility (i.e., hypothesis), to end up at reality conceived of as a realized sector of the total number of possible combinations" (Inhelder and Piaget, 1958, p. 323). On the other hand, proportional reasoning requires the subject to cast all possible combinations "in a double-entry table in such a way as to forecast proportions" (Inhelder and Piaget, 1958, p. 314) qualitatively and later quantitatively (i.e., $x/y = x'/y'$).

METHOD

Selection of Formal Operational Tasks

Formal logic requires an individual to use reasoning abilities that lead the subject towards and the evaluation of evidence which supports or rejects hypothetical cause and effect relationships. These logical operations are used in combinatorial reasoning, isolation and the control of variables, probabilistic reasoning, and proportional reasoning. Table 1 identifies each of the twelve tasks chosen for inclusion in this study and the type of formal operational logic required to solve each task. These twelve tasks were selected for administration to the study population to allow a "multiple-view" perspective of the subject's ability to solve problems using propositional logic and three of the eight formal logical schemata. The four tasks after Renner (1977) were selected because of their ease of use and their combined reliability index of .62 in predicting the "true" Piagetian Score on commonly used, orally presented Piagetian problems. The seven tasks after Lawson (1978) were chosen because of their ease of use, their content (i.e., at least two tasks to assess each formal logical schemata and propositional logic were wanted to fulfill the "multiple-view" perspective), and their reliability index of .86.

Task Construction and Administration

In all, twelve Piaget-type tasks were identified or constructed for this study. Seven of the tasks involved a videotape demonstration using some apparatus and/or physical materials and the remaining five tasks involved written protocols consisting of single problems to be solved. All twelve items posed a question or called for a prediction. The study sample responded to each task in writing in individual response booklets. The medical students were instructed to respond to each task by checking the box next to the best answer or by responding directly to the problem statement, and, in either case, were asked to explain why they responded as they did.

Each item was considered correct (i.e., Level IV-B) only if the correct box was checked and included an adequate explanation for the problem solution. A brief description of each of the twelve Piaget-type tasks follows:

TASK 1: SWINGERS: Using three pendulums (two 18 inches long but with bobs of 50-units and 100 units of mass, the third 24 inches long with a 50-unit bob), the students were asked to identify which of the pendulums should be used in an experiment to find out if the variable of length effects the period of the pendulum. (After Controlling Variables-1, Lawson, 1978).

TASK 2: MORE SWINGERS: Using the same three pendulums as in Task 2, the students were asked to select which pendulums should be used in an experiment to find out if the weight of the bobs effects the period of the pendulum. (After Controlling Variables-2, Lawson, 1978).

TASK 3: BALANCING: Given a balance beam and hanging masses, the students were asked to predict where a 5-unit mass should be hung to balance a 10-unit mass which is hung 7 units of length from the fulcrum. (After Proportional Reasoning-3, Lawson, 1978).

TASK 4: MORE BALANCING: Using the same balance beam, the students were asked to predict where a 10-unit mass should be hung to balance a 15-unit mass which is hung 4 units of length from the fulcrum. (After Proportional Reasoning-4, Lawson, 1978).

TASK 5: SWITCHES: Given a wooden box with five color-coded switches and a light with an associated activation button, the students are shown that the light can be activated by flipping a certain combination of switches and pushing the activation button. They are then asked to list all of the possible combinations of the five switches that they would have to try to discover which combination or combinations will activate the light. (After Switches, Hale, 1972 and 1976 and Combinatorial Reasoning-1, Lawson, 1978).

TASK 6: GRAB BAG: Three red cubes, four yellow cubes, and five blue cubes were placed in a hat. Four red spheres, two yellow spheres and three blue spheres are also placed in the hat. The students are asked to predict the chances of drawing out a red piece on the first try. (After Probability-2, Lawson, 1978).

TASK 7: MORE GRAB BAG; Using the same wooden pieces as in Task 7, the students are asked to predict the chances of drawing a red or a blue sphere on the first try. (After Probability-3, Lawson, 1978).

TASK 8: FROG COUNT: Given that an ecologist caught and banded 55 frogs and then returned them to the pond to redistribute themselves, and given that the ecologist returned a week later and caught 75 frogs, 12 of which had bands, the students were to estimate from this information the "total" number of frogs in the pond. (After The Frog Problem, Renner, 1977).

TASK 9: SHADOWS: Given that the shadow of a building was determined to be 50 meters and height and shadow of a post was determined to be 3 meters and 2 meters respectively, when both are measured at the same time during a given day, the students were asked to estimate the height of the building. (After The Shadows Problem, Renner, 1977).

TASK 10: ROCK WEIGHING: An 8-pound rock suspended from a scale into a container of water was measured at 6 pounds. Given this information, the students are asked to explain this finding. (After The Rock and Scale Problem, Renner, 1977).

TASK 11: GERANIUMS: The students were asked to describe how they would design an experiment given an unlimited supply of like geranium plants to determine the effects of various amounts of sunlight, fertilizer, and water on their growth. (After The Geranium Problem, Renner, 1977).

TASK 12: SEMINAR: The students were shown a seating chart with 12 seats arranged in three rows, four seats in a row. The students were then asked to determine the probability of two individuals being randomly assigned to any two seats next to one another in any of the three rows. (Original, 1980).

Subjects

Fifty-nine of sixty-five second year medical students at the University of South Dakota agreed to participate in this study. This study group consisted of forty-nine males and ten females, age 21 to 35 years of age (mean age = 23). Each subject was given a Problem-Solving Analysis Study Answer Booklet to record the results of the seven videotape demonstrations (Tasks 1-7) and the results of the five written problems (Tasks 8-12). The twelve tasks were later scored for each individual on a three point system per task. This scoring system is shown in Figure 1.

Points	Piagetian Level	Description
3	Formal (III-B)	A correct logical explanation for a given problem was provided; using a systematic approach.
2	Transitional (III-A)	An incomplete logical explanation for a given problem was provided; using a systematic approach.
1	Concrete (II-A, B)	An attempt was made to solve the problems, but without a systematic approach.
0	Pre Concrete (I)	No attempt was made to solve the problem or the attempt was not relative to the problem presented.

FIGURE 1: Problem Solving Analysis Study Scoring System

RESULTS

The results of the data analysis are presented in Tables 2, 3, and 4. Table 2 describes the number of subjects falling into each of the scoring categories for the twelve Piaget-related tasks.

Table 3 describes the number of subjects falling into each of the scoring categories for the Propositional logic tasks (1, 2, and 11), the combinatorial reasoning tasks (5 and 10), the probabilistic reasoning tasks (6, 7 and 12), and the proportional reasoning tasks (3, 4, 8 and 9).

Table 4 represents the number of subjects whose total score in a logical category represents formal reasoning, transitional formal reasoning and concrete operational reasoning for that given category.

DISCUSSION

Descriptive analysis of the Piaget-related task data indicate the following conclusions:

1. Two (2) of the 59 subjects (sophomore medical students) were rated formal on <u>all</u> 12 tasks (i.e., a total score of 36).

2. Fifty-seven (57) of the 59 subjects were rated transitional formal on the 12 tasks (i.e., a total score of 19-35).

3. No subjects were rated as concrete operational on the 12 tasks (i.e., a total score of less than 19).

4. Twenty-five (25) subjects were rated as formal on the Propositional Logic Tasks; all but one (1) of the remaining were rated transitional formal.

5. Nineteen (19) subjects were rated as formal on the Combinatorial Reasoning Tasks; the remaining were rated as transitional formal.

6. Nine (9) subjects were rated as formal on the Probabilistic Reasoning Tasks; the remaining were rated as transitional formal.

7. Forty-five (45) subjects were rated as formal on the Proportional Reasoning Tasks; the remaining were rated as transitional formal.

8. Fifty-three (53) subjects were rated as formal on the 12 tasks (i.e., a total score of 31 or better on all 12 tasks) if modified scoring criteria are used to rate them (i.e., subjects obtaining a majority of scores of 3 on the twelve tasks).

9. The depression of some scores for the group (i.e., for tasks 5, 10, 11, and 12) could be accounted for because of vague directions provided in the videotape and/or response booklet. This researcher provided verbal clarification for each of these tasks during the testing session. Task 12, "Seminar", was a particularly difficult probability problem.

10. Given that from 16-80% of college freshmen and/or sophomores (Renner, 1979; Killian, 1979; Schwebel, 1975; and Lawson, Karplus, and Adi, 1978) and from 43-77% of adults--non-college and college graduates age 20-60 years (Sennott, 1975; Kohlberg and Gilligan, 1971: and Tomlinson-Keasey, 1972), the subjects of this study (i.e., fifty-three of which were classified as either transition or formal operational on all tasks) comprised ninety percent (90%) formal level thinkers on these twelve tasks.

CONCLUSIONS AND IMPLICATIONS

Of the fifty-nine (59) second year medical students who participated in this study, two can be considered fully formal on the twelve Piaget-related tasks used in this study. Of the remaining fifty-seven (57) subjects, all were rated as transitional formal on these tasks (scores of 19-35). Given that fifty-three (53) of this group scored at the 31-35 point level on the tasks, and given that the remaining subjects exceeded a score of 19 on all tasks combined, it is quite possible that their transition to fully formal operational reasoning is nearly complete or complete since testing error could account for the loss of these few points in the overall testing process. Given the hypothetico-deductive ability that Piaget states is required for complex problem solving, this particular select population of adults seems well trained and/or developed intellectually to meet the demands of their medical education and their future roles as medical problem solvers (i.e., physicians).

Future study should better define the characteristics of medical students in terms of their level of logical thinking. Such areas as MCAT scores, NBME, Part I scores, Grade Point Averages, grades in undergraduate science, grades in medical bioscience courses, and personal characteristics will be viewed in relation to these measures of problem solving ability (i.e., Piagetian Tasks). It is anticipated that such future studies will identify additional characteristics of medical student populations which may be useful for student selection, advisement, and training.

REFERENCES

Hale, J., "An Investigation of Two Formal Operational Schemata in Adolescents in the ISCS Classrooms of Three Selected Teachers", Unpublished Doctoral Thesis, The University of Iowa, Iowa City, Iowa, 1972.

Hale, J., "An Investigation of Two Formal Operational Schemata in Adolescents Enrolled in ISCS Classrooms", Journal of Research in Science Teaching, 1976, Vol. 13 (6), pp. 559-564.

Inhelder, B. and J. Piaget, The Growth of Logical Thinking: From Childhood to Adolescence, 1958, Basic Books, Inc.

Killian, C., "Cognitive Development of College Freshmen", Journal of Research in Science Teaching, 1979, Vol. 16 (4), pp. 347-350.

Kohlberg, L. and C. Gilligan, Daedalus 100, Vol. 1051, Fall 1971.

Lawson, A., "The Development and Validation of a Classroom Test of Formal Reasoning", Journal of Research in Science Teaching, 1978, Vol. 15 (1), pp. 11-24.

Lawson, A. R. Karplus and H. Adi, "The Acquisition of Propositional Logic and Formal Operational Schemata During the Secondary School Year", Journal of Research in Science Teaching, 1978, Vol. 15 (6), pp. 465-478.

Renner, J., "The Relationship Between Intellectual Development and Written Responses to Science Questions", Journal of Research in Science Teaching, 1979, Vol. 16 (4), pp. 279-299.

Renner, J., D. Prickett and M. Renner, Evaluating Intellectual Development Using Written Responses to Selected Science Problems, A Report to the National Science Foundation on Grant No. EPP75-19596, 1977, The University of Oklahoma.

Schwebel, M., "Formal Operations in College Freshmen", Journal of Psychology, 1975, Vol. 91, pp. 133-141.

Sinnott, J., "Everyday Thinking and Piagetian Operativity in Adults", Human Development, 1975, Vol. 18 (6), pp. 430-443.

Tomlinson-Keasey, C., "Formal Operations in Females from Eleven to Fifty-four Years of Age", Developmental Psychology, 1972, Vol. 6 (2), page 364.

TABLE 1

THE TWELVE PIAGET-RELATED STUDY TASKS

A. Isolation and Control of Variables (Propositional Logic)

1. "Swingers" (after Lawson, 1978)
2. "More Swingers" (after Lawson, 1978)
3. "Geraniums" (after Renner, 1977)

B. Combinatorial Reasoning (Formal Logical Schemata I)

1. "Switches" (after Hale, 1972 and Lawson, 1978)
2. "Rock Weighing" (after Renner, 1977)

C. Probabilistic Reasoning (Formal Logical Schemata II)

1. "Grab Bag" (after Lawson, 1978)
2. "More Grab Bag" (after Lawson, 1978)
3. "Seminar"

D. Proportional Reasoning (Formal Logical Schemata III)

1. "Balancing" (after Lawson, 1978)
2. "More Balancing" (after Lawson, 1978)
3. "Frog Count" (after Renner, 1977)
4. "Shadows" (after Renner, 1977)

TABLE 2

PIAGET RELATED TASK SCORES BY TASK

	Title	Scores 1	2	3
Task 1:	"Swingers"	3	1	55
Task 2:	"More Swingers"	2	0	57
Task 3:	"Balancing"	1	1	57
Task 4:	"More Balancing"	2	2	55
Task 5:	"Switches"	2	20	37
Task 6:	"Grab Bag"	0	0	59
Task 7:	"More Grab Bag"	1	5	53
Task 8:	"Frog Count"	3	7	49
Task 9:	"Shadows"	0	1	58
Task 10:	"Rock Weighing"	3	19	37
Task 11:	"Geraniums"	0	32	27
Task 12:	"Seminar"	16	32	11

*No "0" scores were obtained by the 59 subjects on any of the twelve tasks.

245

TABLE 3
PIAGET-RELATED TASK SCORES
BY LOGICAL CATEGORY

Logical Category	Score 1	2	3
1. Propositional Logic 3 Tasks x 59 = (177 possible)	5	33	139
2. Combinatorial Reasoning 2 Tasks x 59 = (118 possible)	5	39	74
3. Probabilistic Reasoning 3 Tasks x 59 = (177 possible)	17	37	123
4. Proportional Reasoning 4 Tasks x 59 = (236 possible)	6	11	219

TABLE 4
TOTAL PIAGET-RELATED TASK SCORES
BY LOGICAL CATEGORY

Logical Category	Scores: Concrete	Transitional	Formal
1. Propositional Logic (3 Tasks)	1 (Range=3-4 points)	33[1] (Range=5-8 points)	25 (9 points)
2. Combinatorial Reasoning (2 Tasks)	0 (2 points)	40[2] (Range=3-5 points)	19 (6 points)
3. Probabilistic Reasoning (3 Tasks)	0 (Range=3-4 points)	50[3] (Range=5-8 points)	9 (9 points)
4. Proportional Reasoning (4 Tasks)	0 (Range=4-5 points)	14[4] (Range=6-11 points)	45 (12 points)
TOTAL (12 Tasks)	0 (Range=12-18 points)	57[5] (Range=19-35 points)	2 (36 points)
$\overline{X}$=32.84	S.D.=2.02		

1. Thirty-one (31) of the 33 had a score of 8 on this task.
2. Thirty (30) of the 40 had a score of 5 on this task.
3. Thirty-three (33) of the 50 had a score of 8 on this task.
4. Thirteen (13) of the 14 had scores of 10 or 11 on this task.
5. Fifty-one (51) of the 57 had scores of 31 to 35 on these tasks (a score of 30 or above would be considered formal by some researchers).

PRECIS

PREDICTABILITY OF PREDICTIVE TECHNIQUES

MODERATOR: Roger Girard, Ph.D.
University of Southern California, School of Medicine

PATH ANAYLSIS OF MEDICAL STUDENT PERFORMANCE DATA

Measures of medical student performance from pre-admission through intership are examined using path analysis. The intent is to clarify the complex interrelationships that customarily exist between such variables.

CANONICAL REDUNDANCY ANALYSIS: A NEW TECHNIQUE TO PREDICT PERFORMANCE

A new statistical method, canonical redundancy analysis, is used to supplement the traditional univariate analyses in predicting clinical performance in medical school. Overall 16 percent of the variance in all clinical measures is accounted for, with the undergraduate nonscience grade point average being the key predictor variable.

INCREMENTAL VALIDITY: THE OLD AND NEW MCATs COMPARED

Incremental validity is the extent to which a test will increase the accuracy of predictions made on the basis of data that are usually available. This study compares the incremental validities of the Old and New MCATs in predicting medical student performance in the first two years.

THE RELATIONSHIP BETWEEN MCAT SCIENCE SUBTEST SCORES AND PERFORMANCE IN MEDICAL SCHOOL - THE IMPACT OF THE UNDERGRADUATE INSTITUTION

The predictive validity of the MCAT with regard to medical students' performance in medical school appears to be, in part, a function of the undergraduate college the students have attended.

RESTRICTION OF RANGE AND THE PREDICTIVE VALIDITY OF THE NEW MEDICAL COLLEGE ADMISSION TEST

Predictive validity coefficients for the New Medical College Admission Test were corrected for restriction of range in a reliable criterion (.92). It was found that such corrections substantially increases the validity coefficients.

Path Analysis of Medical Student Performance Data*

Charles P. Friedman, Ph.D.
University of North Carolina at Chapel Hill

Introduction

Examination of extensive literature on the prediction of medical student performance suggests that investigators have largely overlooked a very potent and appropriate analytic tool.[1] The method of path anlysis, first introduced by Sewall Wright in 1934, would appear to be ideally suited to treatment of the multivariate data sets endemic to this area of research.[2] Computationally, path analysis may be viewed as a straightforward extension of multiple linear regression techniques; however, there is a major underlying difference. Use of path analysis requires a pre-statement by the researcher of a theoretical model of posited causal interrelationships between variables of interest. Subsequent quantitative analysis closely related to multiple regression then provides a measure of the goodness-of-fit between theory and data. Path analytic approaches can also allow for a more efficient comparison, dissemination, and generalization of research results, because these results are based on conceptual models that can be expressed diagrammatically and readily interpreted.

This paper is very much in the tradition of those attempting to examine the interrelationship between medical student performance variables over a period spanning pre-admission and post-graduate training.[3-5] These studies have largely employed correlational or multiple linear regression methods; as, to this author's knowledge, only one unpublished manuscript has addressed a similar task using path analysis.[6] To the extent that one may generalize about such studies, they have tended to show that the best predictor of a given year's performance is performance in the previous year, and that ratings of clinical skills are not well predicted by either pre-admission data or by measures of achievement in the basic sciences. Path analytic methods can bring added sharpness to findings of this type. Some studies have demonstrated the unique contribution to prediction that non-cognitive factors can provide.[5,7] The present study will not consider such non-cognitive variables, but will demonstrate how such measures could be readily blended into the overall analytical framework that is employed.

Variables and Subjects

The study examines the performance of two classes of students at the University of North Carolina (UNC) School of Medicine: those entering in 1973 and 1974. These are the most recent classes for which full sets of longitudinal data are available. The measures themselves derive from data that are routinely available as products of the student evaluation system. The variables included in the model to be tested are:

- 8 pre-admission measures (four "Old" MCAT scores, undergraduate science and non-science GPA, an index of "quantitativeness" of undergraduate major, and an index of selectivity of each student's undergraduate school computed from the school's mean SAT scores);

*Reprints available from: Office of Medical Studies, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27514

- two basic science performance measures (scores on first and second year medical school comprehensive examinations);
- two grade point averages, one for third year clinical clerkships and the other for fourth year electives;
- total score on a comprehensive examination (comprising five NBME part II sub-tests) taken at the end of the third year;
- ratings of each student's performance as an intern by the student's internship supervisor. (The actual measure used is the mean response to 11 items on the evaluation form.)

Because they require consideration of all pairs of variables in a model, path analyses quickly become overwhelmingly complex as the number of variables increases. For this reason, the eight pre-admission variables were factor analyzed to produce a more parsimonious set. The resultant was a two factor solution (Table 1), with one factor loading on the MCAT and undergraduate school selectivity, and the other factor loading on undergraduate GPAs. For analytical purposes, composite indices on each factor were constructed by summing the variables loading heavily on that factor. While computationally straightforward, this method raises the possibility that the factor scores themselves will be moderately inter-correlated in the study sample.

Analyses were performed using only those subjects with no missing data (N=167). The analytical sample comprises 76% of the total population of the two classes under study. Most of the exclusions were the result of missing internship ratings; however, the mean third year examination scores and the third and fourth year grade point averages were examined and shown to be homogeneous for students with and without internship ratings. A more substantive but less numerous omission from the analytical sample comprises those students who did not complete the curriculum (<10); thus it must be emphasized that the results reported here apply to students who entered the UNC School of Medicine and went on to post-graduate training.

Causal Model and Analytic Approach

The first step in the path analysis process is specification of a theoretical model which indicates all posited causal linkages between the variables included.

Testing or estimaton of a path analytic model requires multiple linear regression of each variable in the model on all variables specified as causally prior to it.* The resulting standardized regression coefficients provide the best estimate of the unique direct effect of one variable on another for each relationship specified. The coefficients are called path coefficients and appear alongside an "arrow" in a path diagram to indicate the magnitude of a computed direct effect. The goodness-of-fit between the model and data is determined using Wright's algorithm to calculate, based on the structure of the theoretical model and the magnitudes of the computed path coefficients, a "reproduced" correlation matrix. If the model specified is a good fit, the

*This statement applies only to the class of models that may be termed "recursive."

reproduced correlation matrix will closely resemble the observed correlation matrix from which the path coefficients were originally computed.

In the absence of a comprehensive theory to inform generation of a conceptual model of student performance, the initial step in data anlaysis for this study was based on the least-parsimonious recursive model. That is, the initial model tested posited a direct causal relationship between each performance measure and all variables assessed at a later point in curriculum time. The path coefficients for this model were estimated and weak relationships removed. The resultant "trimmed" model was then reestimated--the path coefficients recomputed--and the goodness-of-fit of this model to the data was determined.

Results

Figure 1 gives the trimmed path model with its path coefficients and Table 2 gives the matrix of observed and reproduced correlations. In Figure 1 it should be noted that the correlation between the two pre-admission factor scores was specified from the observed correlation matrix; it is a "given" in the estimation process. One may infer from Figure 1 the large number of relationships between variable pairs that were omitted via the trimming process. A relationship was omitted if the magnitude of the relevant path coefficient was less than .05 or, if greater than .05, it was not statistically significant with $p < .05$. For example, in the estimation of the original model, fourth year grade point average was regressed on both pre-admission factors, first, second, and third year comprehensive examinations, and third year GPA. Only two variables--third year GPA and third year comprehensive exam score--proved sufficiently significant to survive the trimming process. Also shown in Figure 1 are the disturbance paths, representing the strength of the direct relationship between each variable and all of the unspecified factors that account for the "unexplained" variance in that variable.

From Table 2 it may be inferred that the reproduced correlation matrix is in high conformity with the matrix of observed correlations computed with the raw data. The root-mean-square deviation between reproduced and observed correlations is .03; and this calculation includes only correlations which are not constrained to be equal in both matrices by the structure of the model. Since the average magnitude of the intercorrelations in the observed matrix is .43, the error in reproducing the correlation matrix is 6.9%.

Discussion

Based on the concordance of the observed and reproduced matrices, we conclude that the trimmed model in Figure 1 is descriptive of the interrelationship between the variables included. The causal "chain" structure for successive years of performance is markedly in evidence, lending further weight to the contention that most recent past student performance is the best predictor of present student performance. The one significant exception to this, observable in Figure 1, is the moderate relationship between internship ratings and third year grade point average. The result may be attributable to the structure of the UNC curriculum wherein the fourth year is highly elective and diverse, while the third year emphasizes clerkships in inpatient services similar to the work setting of most interns. While a substantial correlation is observed between pre-admission measures and test performance in the third medical school year, the model shows this to be an indirect effect, with earlier medical school performance as an intervening variable.

Another effect evident from the magnitude of the path coefficients is a tendency for measures of the same type to be strongly related. Examinations predict other examinations; and, if grades are viewed as a type of rating, ratings may be said to predict other ratings. The strong direct relation between second year and third year examination scores indicates that, despite the difference in content domain, students who perform well in cognitive tests of basic science knowledge also perform well in similarly constructed tests of clinical knowledge.

If internship ratings are, in a theoretical and substantive sense, the ultimate outcome variable of this study, then the magnitude of the disturbance path to this variable is a strong reminder of the level of uncertainty involved in ranking students for internship with heavy weight on their performance in medical school. Ratings by supervisors may not be a totally valid performance measure, but clearly such ratings are important. Other factors, very possibly some non-cognitive factors employed in studies cited earlier, may be necessary to provide a clearer picture of the causes of highly rated clinical performance after graduation.

References

1. Cuca, J.M., Sakakenny, L.A., and Johnson, D.G. The Medical School Admission Process: Review of the Literature 1955-76. Washington: Association of American Medical Colleges, 1976.

2. Duncan, O.D. Introduction to Structural Equation Models. New York: Academic Press, 1975.

3. Korman, M., and Stubblefield, R.L. Medical School Evaluation and Internship Performance. J. Med. Educ., 46: 670-673, 1971.

4. Veloski, J., Herman, M.W., et al. Relationships Between Performance In Medical School and First Postgraduate Year. J. Med. Educ., 54: 909-916, 1979.

5. Turner, E.V., Helper, M.M., and Kriska, S.D. Predictors of Clinical Performance. J. Med. Educ., 49: 338-342, 1974.

6. Wysocki, J.C. A Mathematical Model to Examine Background Variables and Events Which Predict Performance on NBME Tests. Technical Report of the Office of Medical Education, University of Arizona, 1977.

7. Roessler, R., Lester, J.W., et al. Cognitive and Noncognitive Variables in the Prediction of Preclinical Performance. J. Med. Educ., 53: 678-680, 1978.

TABLE 1: Varimax Rotated Factor Matrix for Pre-admission Variables

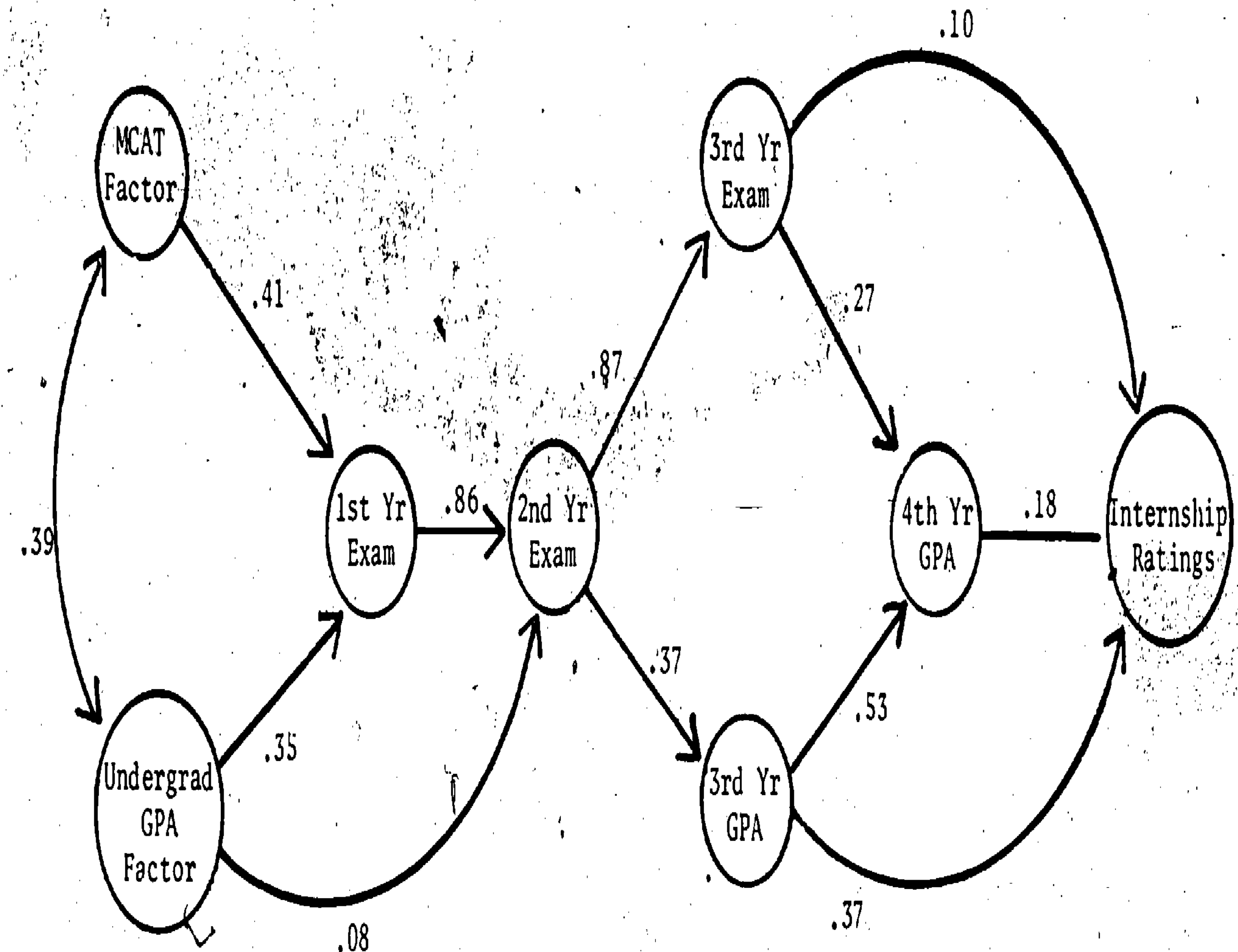
Variable	Factor 1 (MCAT) Loading	Factor 2 (GPA) Loading	Communality
Quantitative MCAT	.609	.299	.460
General Information MCAT	.743	.221	.600
Verbal MCAT	.842	.158	.735
Science MCAT	.682	.283	.545
Science GPA	.136	.987	.992
Non-Science GPA	.145	.625	.412
Quantitativeness Index of Major	-.157	.102	.035
Undergraduate School Selectivity	.616	-.013	.379
Eigenvalue	3.07	1.09	

TABLE 2: Observed and Reproduced Correlations

(Reproduced / Observed)

	MCAT Factor	Undergrad GPA Factor	1st Yr Exam	2nd Yr Exam	3rd Yr Exam	3rd Yr GPA	4th Yr GPA	Ratings
MCAT Factor		.39	.55	.50	.42	.22	.24	.17
Undergrad GPA Factor	.39		.51	.52	.42	.24	.29	.22
Exam 1	.55	.50		.90	.82	.35	.39	.31
Exam 2	.50	.52	.90		.87	.37	.44	.26
Exam 3	.44	.45	.79	.87		.37	.47	.32
GPA 3	.19	.19	.33	.37	.32		.63	.52
GPA 4	.21	.22	.39	.43	.44	.62		.46
Intern Ratings	.15	.16	.27	.30	.30	.51	.45	

FIGURE 1: Path Diagram for "Trimmed" Model



Disturbance Paths

To:	Magnitude
1st Yr Exam	.72
2nd Yr Exam	.48
3rd Yr Exam	.49
3rd Yr GPA	.93
4th Yr GPA	.74
Internship Ratings	.83

-244-

Canonical Redundancy Analysis: A New Technique to Predict Performance

Beth Dawson-Saunders and Deane R. Doolen
Southern Illinois University School of Medicine*

Medical educators continue to be interested in the relationships between indicators of performance and actual measures of performance. The capability to predict performance is needed to make the best informed admission decisions as well as to identify and plan remedial instruction for students with predictable weaknesses.

The majority of research studies on relationships between variables related to medical school performance report only correlations and/or multiple regression analyses (1-6). The correlations between given predictor variables and various performance measures are often inconsistent, and it is generally quite difficult to evaluate the overall relationship between the two sets of variables. The canonical redundancy statistic was developed in 1968 to measure the entire amount of variance in one set of variables accounted for by the other set of variables (7). In this sense, it is a multivariate analogue to the coefficient of determination: r^2 in correlation and R^2 in multiple regression. In fact, the redundancy statistic is equal to the average R^2 obtained when all variables in one set are regressed on each variable in the second set.

Redundancy analysis appears to be an ideal technique to supplement the traditional univariate analyses used in medical education research. Thus the purpose of the present paper is to evaluate the usefulness of canonical redundancy analysis in examining the relationship between nine pre-selection variables and four measures of clinical performance in medical school.

METHOD

Subjects were students in the 1979 and 1980 graduating classes of the Southern Illinois University School of Medicine. The three year program includes 18 months of basic science instruction, using an integrated organ system approach, and 18 months of clinical instruction. A total of 143 students were included in the analysis; four advanced standing students were eliminated from the sample since the criteria for their acceptance were atypical from that of traditional entering students.

Pre-selection Characteristics. Nine characteristics of applicants were used as the predictor set of variables. These included the four old MCAT scores and cumulative undergraduate science (BCPM) and nonscience (AO) grade point averages. Values of these variables were taken from the AMCAS application data card. Additionally considered were the year of birth, sex and a measure of traditional versus nontraditional undergraduate background. A student's background was considered traditional if it included a science or "pre-med" major and if the student, as an undergraduate, either planned or applied to medical school. A nontraditional background was a nonscience major or a career change before applying to medical school.

*Reprints available from Beth Dawson-Saunders, Southern Illinois University School of Medicine, P.O. Box 3926, Springfield, IL 62708

Clinical Performance Characteristics. Students from SIU School of Medicine experience six major clinical rotations: family practice, medicine, obstetrics and gynecology, pediatrics, psychiatry and surgery. At the completion of each rotation, faculty rate and describe each student's abilities on the characteristics of clinical sophistication, cognitive knowledge, personal maturity and communication skills. Clinical sophistication includes abilities in history taking, physical examination, problem identification and integration, use of tests and procedures and patient management. Cognitive knowledge is defined as a student's general fund of knowledge and independent learning abilities. Personal maturity consists of reliability and dependability, ability to assume responsibility, maturity and relationships with patients, their families and other medical personnel. Communication skills involves both written and oral abilities.

The ratings and descriptions are considered as primary information in writing letters of recommendation for residency programs. Faculty rate each student as exceptional, satisfactory or unsatisfactory on each of the four clinical performance characteristics. Additionally, for purposes of the present study, the written descriptions were content analyzed to form three categories from the satisfactory ratings. Thus, ratings for each student were as follows: 5=exceptional, 4=very good, 3=satisfactory, 2=marginal and 1=unsatisfactory.

Statistical Procedure. Correlations between the six clinical rotations were computed on each of the four clinical performance measures. Seventy-seven percent of the correlations were greater than .14 ($p<.05$); therefore, the ratings on each clinical measure were summed across the six rotations, forming a scale more appropriate for multivariate analysis.

Correlations and multiple regression analyses were done to determine the univariate relationships between the pre-selection variables and each clinical performance measure. Canonical correlation and redundancy analysis was performed on the two sets of variables to determine the overall relationship between them. While the redundancy statistics may be computed from the multiple regression results, computer programs have been developed to compute their values directly from the results of canonical correlation analysis (8,9). The factor structure of correlations between the pre-selection and clinical variables and the canonical variates were computed to facilitate interpretation of their relationships (10).

RESULTS

The zero order correlations between the pre-selection variables and clinical performance measures are shown in Table 1. Overall, the correlations are positive and low to moderate in size. The nonscience GPA is the only variable significantly correlated with all clinical measures. Science and quantitative MCAT's are significantly correlated only with clinical sophistication and cognitive knowledge. The science GPA is also related to these two measures and weakly to communication skills. Year of birth is a correlate of cognitive knowledge, indicating higher ratings for younger students. Sex is related to personal maturity and communication skills, with females rated higher. Finally, the verbal MCAT is weakly related to communication skills.

The results of the multiple regression analyses are given in Table 2. As indicated by the correlations in Table 1, cognitive knowledge is best predicted with an F-ratio of 4.33 (df=9,133; $p<.001$) and a multiple correlation of R=.48. Clinical sophistication and communication skills are predicted with F-ratios of 3.00 and 2.69 (df=8,134; $p<.01$) and multiple correlations of .39 and .37, respectively. The regression on personal maturity has an F-ratio of 2.93 (df=6,136; $p<.05$) and R=.34.

The squared canonical correlations decrease from .26 for the first variate pair to .05 for the fourth (Table 3). The redundancy statistics are similar in size for the first variate, .07 and .09 respectively, but the redundancy for the clinical performance measures remains at .06 for the second variate also. Overall, the total redundancy for the clinical measures, given the predictor variables, is .16, indicating that 16 percent of the variance in clinical performance may be accounted for by knowledge of the pre-selection variables. A check of Table 2 show that this is indeed the average R^2 obtained in the four multiple regression analyses. The corresponding redundancy for the pre-selection measures is .10 and is, of course, of less interest since the desire is to predict in the other direction.

The factor structures are presented in Table 4. The first variate pair, accounting for 56 percent of the redundancy in clinical performance, is primarily a GPA and science MCAT relationship with clinical sophistication and cognitive knowledge, with lesser contributions by the general information and quantitative MCAT's. The second variate pair, accounting for 38 percent of the redundancy, appears to be a nontechnical dimension: verbal MCAT and nonscience GPA with personal maturity and communication skills. The third and fourth variates contribute so little to the redundancy that it is best not to interpret them.

DISCUSSION

In general, the results confirm that of other reported attempts to predict clinical performance: there are positive but overall weak to moderate relationships between pre-selection and clinical measures of performance. The variable most related is the undergraduate nonscience GPA, indicating perhaps a more general ability required in clinical achievement than that measured by only science abilities. It must also be noted that the science GPA and MCAT are frequently accorded greater strength in the selection process, resulting in unavoidable restriction of range problems in evaluating these measures.

The science and quantitative MCAT's, as well as the undergraduate science GPA, are significantly related to clinical sophistication and cognitive knowledge. However, they provide little assistance in an attempt to predict ratings of maturity and communication. Here the verbal MCAT is a strong indicator, based on the canonical analysis. The relationship with sex is weaker when other variables are included in the multivariate analysis; it may be that the role of sex is moderated when considered with the verbal MCAT since females tend to score higher on this component.

The fact that age is related to cognitive knowledge on the basis of the univariate analysis is somewhat interesting: younger students tend to be rated higher on this measure. A possible explanation is that younger students are often those selected the first year of their application. However, the relationship is much weaker on the canonical analysis, perhaps again due to a moderating effect by the other variables.

The results of the redundancy analysis soberly attest to our limited ability to predict clinical performance on the basis of pre-selection variables. Of course, there is at least a three-year span between the measurements, and correlations generally decrease over such a time period. Secondly, the pre-selection variables are primarily indicators of cognitive ability while clinical performance, even in the cognitive realm, is a more complex construct. In the present research, the clinical performance measures were the result of faculty ratings and, therefore, may not be as objectively based as the pre-selection variables. However, the nature of clinical evaluation generally involves a rating-type method and this reality must be recognized. It may be that the new MCAT examinations will result in measures more closely associated with clinical performance; this research must wait another year or two.

While the redundancy statistics are low, they do appear to provide a more realistic evaluation of our present ability to predict clinical performance. Canonical and redundancy analysis is, of course, subject to the same vagaries that haunt all correlational methods; it does, however, provide a method to examine the overall relationship between two sets of variables and one that is consistent with the univariate procedures.

CONCLUSION

Pre-selection variables were found to be weakly related to ratings of clinical performance for two classes of medical students. The variable most related was undergraduate nonscience grade point average. Other significant predictor variables included the science GPA and the science, quantitative and verbal MCAT's. Sex and age of the student were weak predictors that appear to be moderated when combined with the other variables. The use of canonical redundancy analysis provides an evaluation of the overall relationship of these variables to clinical performance consistent, but not generally possible, with the usual method of examining numerous zero-order correlations and regression equations.

REFERENCES

1. Turner, E.V., Helper, M.M. and Kriska, D.S. Predictors of Clinical Performance. J. Med. Educ., 48:338-342, 1974.

2. Deighton, B.C., Smith, D.U., and Gallagher, R.E. The Relationships Between Scores on the New MCAT and Performance in the First Year of Medical School. RIME Proc.:133-137, 1979.

3. Willoughby, T.L., Gammon, L.C., and Jonas, H.S. Correlates of Clinical Performance During Medical School. J. Med. Educ., 54:453-460, 1979.

4. Shaw, D.L. Randels, P.M. and Elson, P. Does MCAT Predict Clinical Performance? J. So. Carolina Med. Assn., 75:371-374, 1979.

5. Murden, R., Galloway, G.M., Reid, J.C. and Colwill, J.M. Academic and Personal Predictors of Clinical Success in Medical School. J. Med. Educ., 53:711-719, 1978.

6. Kegel-Flom, P. Predicting Supervisor, Peer, and Self Ratings of Intern Performance. J. Med. Educ., 50:812-815, 1975.

7. Stewart, D. and Love, W. A General Canonical Correlation Index. Psych. Bull., 70:160-163, 1968.

8. Cooley, W.W. and Lohnes, P.R. Multivariate Data Analysis. New York: Wiley, 1971.

9. Thompson, B. and Frankiewicz, R.G. Canon: A Program Which Produces Canonical Structure and Index Coefficients. Educ. and Psych. Meas., 39:319-322, 1979.

10. Tatsuoka, M.M. Multivariate Analysis. New York: Wiley, 1971.

TABLE 1: Correlations Between Pre-selection Characteristics and Performance on Clinical Rotations

Pre-selection Characteristics	Clinical Performance Characteristics			
	Clin Soph	Cog Know	Pers Mat	Comm Skills
Year of Birth	.12	.25***	.07	.05
Sex(1)	.07	.02	.19*	.20**
Undergrad. Bkground(2)	.01	-.09	.03	.10
Gen. Info. MCAT	.08	.16*	.00	.11
Quantitative MCAT	.25**	.21**	.03	.13
Verbal MCAT	.12	.11	.05	.18*
Science MCAT	.18*	.21**	.00	.13
Science GPA (BCPM)	.27***	.39***	.13	.16*
Nonscience GPA (AO)	.34***	.38***	.27***	.24**
Mean±sd	22.0±2.6	21.5±2.6	23.5±2.6	22.0±2.0

(1) 1=male, 2=female (2) 1=traditional, 2=nontraditional
*p<.05; **p<.01; ***p<.001

TABLE 2: Multiple Regression on Clinical Performance Characteristics

Clinical Performance Characteristic	Coeff of Mult Corr (R)	Variables Not in Equation	F	df
Clinical Sophistication	.39	Gen. Info. MCAT	3.00**	8,134
Cognitive Knowledge	.48		4.33***	9,133
Personal Maturity	.34	Gen. Info. MCAT Verbal MCAT Science GPA	2.93*	6,136
Communication Skills	.37	Science GPA	2.69**	8,134

*p<.05; **p<.01; ***p<.001

TABLE 3: Squared Canonical Correlations and Redundancy Statistics

Variate Pair	Squared Canonical Correlation (λ^2)	Redundancy Statistics Pre-selection	Redundancy Statistics Clinical
1	.26	.07	.09
2	.13	.02	.06
3	.08	.01	.01
4	.05	.00	.00
Total Redundancy		.10	.16

TABLE 4: Factor Structure Dimensions for Pre-selection and Clinical Performance Characteristics

Pre-selection Characteristics	Variate Loadings 1	2	3	4
Year of Birth	.21	.12	.22	-.24
Sex(1)	-.14	.27	-.07	-.19
Undergrad. Bkground(2)	-.12	-.17	.06	.22
Gen. Info MCAT	.51	.26	.13	.68
Quantitative MCAT	.51	.11	.94	-.21
Verbal MCAT	.35	.56	.31	.60
Science MCAT	.64	.32	.61	.39
Science GPA (BCPM)	.91	.26	.16	.01
Nonscience GPA (AO)	.64	.78	.20	-.40
Clinical Performance Characteristics				
Clinical sophistication	.59	.41	.23	-.63
Cognitive knowledge	.90	.18	-.24	-.28
Personal Maturity	.25	.71	-.46	-.61
Communication skills	.35	.81	-.01	-.02

(1) 1=male, 2=female (2) 1=traditional, 2=nontraditional

The authors wish to acknowledge the assistance of Jeffrey Holzworth who performed the content analysis on the student ratings.

Incremental Validity: The Old and New MCATs Compared*

Charles P. Friedman, Ph.D. and Carol Q. Porter, B.S.
University of North Carolina at Chapel Hill

Introduction

A major design objective of the New Medical College Admission Test (New MCAT) was: "to develop an assessment instrument to measure achievement levels and the expected prerequisites which are generally relevant to the study and practice of medicine."[1] Given this objective, students' scores on the New MCAT would be expected to carry substantial, but by no means perfect, power to predict student performance in medical school. Furthermore, because the creation, administration, and scoring of the MCAT generate costs that are borne by the students who take it, the MCAT might be expected to do more than provide useful information that is to some degree predictive. It might also be expected to provide useful information that is unique. The extent to which a test "will add to or increase the validity of predictions made on the basis of data which are usually available" may be defined as the incremental validity of the instrument.[2] A proposed national "truth in testing" bill (H.R. 4949) would require developers of standardized tests to undertake incremental validity studies to demonstrate the usefulness of their instrument over and above other measures that might be available with lesser expenditure of resources.[3]

Much of the research published to date on the New MCAT has either examined its internal structure, or has employed zero-order correlations to relate MCAT scores to measures of medical student performance. Studies of internal structure have shown that the New MCAT, like its predecessor, is basically a two-factor test.[4] Correlational prediction studies have demonstrated a significant relationship between MCAT scores and performance in medical school.[1,5] The number of schools involved to date in such predictive studies and the consistency of results combine to suggest that the result is generalizable: the New MCAT does have criterion-related validity of a predictive type. In a local study of the New MCAT's incremental validity, adding New MCAT scores to a prediction equation, after all other information routinely available, was shown to increase the accuracy of prediction of first-year medical student performance by the proportional factor of 30%.[6] (That is, the multiple R^2 with the MCAT included was 1.3 times the R^2 without the MCAT.)

This paper extends the incremental validity study cited above, in two ways. First, it includes performance of second year medical students as an additional outcome measure to be predicted. This is now possible because the first medical class admitted with use of the New MCAT has completed two years of study. Second, this paper makes explicit comparison of the incremental validities of the "Old" and New MCAT. As an outgrowth of this comparison,

*Reprints available from: The Office of Medical Studies, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27514.

alternative methods of computing incremental validity are considered. These analyses are undertaken to help admissions committees and others who must interpret MCAT scores understand more fully the similarities of and differences between the old and new versions. In this respect, the present study is conducted in the spirit of investigations of internal structure, comparing the two tests, cited earlier.

Method

Incremental validity is addressed using multiple linear regression techniques. The general approach is to use a stepwise, hierarchical design, adding first in a step-wise procedure all variables available at time of admission with the exception of the MCAT sub-test scores. The MCAT scores are entered only after all other preadmission data have been entered into the equation, to determine how much additional variance in medical student performance can be explained.

The study sample comprises classes of medical students at the University of North Carolina (UNC) entering in 1976 (N=159), 1977 (N=160), and 1978 (N=158). The incremental validity of the Old MCAT is computed using the classes entering in 1976 and 1977, while the class entering in 1978 is employed to compute the incremental validity of the New MCAT. Independent variables include, in addition to MCAT scores, each student's non-science and science GPAs, the selectivity of undergraduate school using the average SAT scores of the school's entering freshman class, marital status, age, the "quantitativeness" of undergraduate major, parental education (in years) and income, hometown community size, average annual change in undergraduate GPA, highest degree earned (expressed in years required to earn it), and total undergraduate science hours.

Dependent variables in the study are end of year examination scores for the first and second years of the UNC medical curriculum. These tests are the sole means by which medical student performance at UNC is formally assessed. Both examinations are administered over three days in the spring of each academic year. The first-year examination comprises six to seven sub-tests, two of which are purchased from the National Board of Medical Examiners (NBME) and the remainder of which are locally produced by the faculty. The second year examination comprises 11-13 sub-tests, three of which are purchased from NBME and the remainder of which are locally produced.

Because test content and format vary slightly from year-to-year, each of the three classes will be considered as a separate analytic group. For each class, the incremental validity of the MCAT will be computed using four different dependent variables: 1) the first year examination total composite score, 2) the total score on the first-year examination sub-tests purchased from NBME, 3) the second-year examination total composite score, and 4) the total score on the second-year examination sub-tests purchased from NBME. Thus, twelve analyses will be performed in all. The first year examination scores are not employed as predictors of second year examination scores, because the intent of this study is to examine the predictive power of admission data and not the interrelationships between performance measures in medical school. Each total composite score is determined by institutional policy; it is a weighted mean of all sub-tests, with weightings based on the curriculum time allocated to the subject matter covered by each sub-test. NBME sub-tests are considered separately from those locally prepared, because previous study of MCAT incremental validity revealed stronger relationships between the MCAT and other nationally standardized tests, than between the MCAT and local exams.[6]

An immediately available measure of MCAT incremental validity is the absolute amount by which R^2 increases when MCAT scores are added last to the prediction equation. In addition, an index of incremental validity is computed by calculating the proportional increase in performance variance explainable by adding MCAT scores last to the prediction equation. This employs the formula:

$$\text{Index 1} = \frac{R^2 \text{ added by MCAT}}{R^2 \text{ for non-MCAT variables only}}$$

A slightly different view of incremental validity generates a second index. The second formula views incremental validity as the proportion of performance variance, initially unaccounted for by the non-MCAT variables, that becomes accounted for when the MCATs are added to the regression equation:

$$\text{Index 2} = \frac{R^2 \text{ added by MCAT}}{1 - (R^2 \text{ for non-MCAT variables only})}$$

Both indices are employed in the analyses and both are required because the R^2 for the non-MCAT variables can vary from class to class as a consequence of fluctuating variance in the predictor variables themselves. Under such circumstances, the two MCAT versions could be shown to differ in incremental validity artifactually, because for some classes there might be more initially unexplained variance available for the MCAT to predict.

Results

Table 1 displays the results of the twelve incremental validity computations: four outcome variables for each of three classes.* The cell entry for each analysis provides the total R^2 of the regression using non-MCAT variables only, the amount by which R^2 is increased by adding MCATs to the equation, the total R^2 of the regression including all variables, and the two indices of incremental validity using the formulae given above. Immediately in evidence from Table 1 is wide variation in the predictive power of admission variables for the different entering classes. The non-MCAT R^2 is lowest for the class entering in 1978 and highest for the class entering in 1976. This fluctuation is likely attributable to fluctuation in the homogeneity of the entering classes, such that the performance of more homogenous classes is more difficult to predict statistically.

*The results for the class entering in 1978 using first year exam scores as an outcome do not agree precisely with results reported earlier by the senior author.[6] This is because the sets of predictor variables used in the two studies differed slightly.

For the first index of incremental validity, Table 1 reveals that the New MCAT (class entering 1978) is clearly superior to the Old (classes entering in 1976 and 1977). For the second index of incremental validity, the results in Table 1 are more complex. The MCAT incremental validities (using index 2) for the classes entering in 1976 (Old MCAT) and 1978 (New MCAT) are comparable, while that for the class entering in 1977 (Old MCAT) is lower.

Table 2 displays the component tests of the Old and New MCAT that contribute predominantly to the incremental validity. Over all incremental validity calculations, two MCAT component tests contributed an average of 89% of the incremental variance that was contributed by the MCAT as a whole. For the class entering in 1976, the old Science component test was most important from an incremental validity perspective; for the following year's class the old General Information component test was roughly equal in importance to the Science test. From calculations using the class entering 1978, the Science Problems and Biology components of the New MCAT are shown to make the highest incremental contribution.

Discussion

The analyses reported here suggest that, for recent classes entering the University of North Carolina School of Medicine, the incremental validity of the New MCAT exceeds that of its older counterpart. On the first index of incremental validity the New MCAT exceeded the Old for both entering classes in the analysis. For the second index, the class admitted with the New MCAT revealed incremental validity at least as high as that for the classes admitted with the Old. These analyses reported here also expand upon two findings reported earlier:[6] that the incremental validity (index 1) for the New MCAT is approximately .30 when second year exam scores are employed as a criterion; and that for the New MCAT one component score makes the majority of the contribution to incremental validity made by the test as a whole.

There are many limitations to the generalizability of these results. First, the study was conducted using the student body of one medical school. Different results may be obtained at different institutions, and collaborative research should be performed to investigate this possibility. Second, concerns are raised by the fluctuation in predictability of performance for the different entering classes in this study. While use of two indices to compute incremental validity corrects for this to some degree, the very existence of the fluctuation suggests that the "Incremental Validity" of the MCAT may not be a universal constant, but rather a function of time. Third, the validity of the criterion measure may be challenged. At the University of North Carolina, however, the measures used are the only indices of performance that are available. Fourth is the ubiquitous problem of restricted range. There is no way of knowing how well the New or Old MCATs would contribute to incremental discriminations between performance of admitted medical students and the (hypothetical) performance of students not admitted.

This study cannot, of itself, answer the global question of whether the new MCAT is a better test than its predecessor or whether the MCAT's validity justifies its existence. The reader is encouraged to examine the data presented here in combination with other data, to formulate a personal or facilitate an institutional conclusion. The authors do, however, strongly endorse viewing the validity and utility of the MCAT from an incremental perspective--examining the unique information contained in the test--in contrast to an approach that employs only simple correlations.

References

1. New Medical College Admission Test: Interpretive Manual. Washington: Association of American Medical Colleges, 1977.

2. Sechrest, L. Incremental Validity: A Recommendation. Educ. and Psych. Meas., 23:153-158, 1963.

3. U.S. House of Representatives. Educational Testing Act of 1979. Bill H.R. 4949, 96th Congress, 1st Session, July 24, 1979.

4. Molidor, J.B., and Elstein, A.S. A Factor Analytic Study of the Old and New MCAT Examinations. In Proceedings of the Eighteenth Annual Conference on Research in Medical Education. Washington: Association of American Medical Colleges, 1979.

5. Deighton, B.C., Smith, D.V., Gallagher, R.E. The Relationship Between Scores on the New Medical College Admission Test and Performance in the First Year of Medical School. In Proceedings of the Eighteenth Annual Conference on Research in Medical Education. Washington: Association of American Medical Colleges, 1979.

6. Friedman, C.P., and Bakewell, W.E. Incremental Validity of the New MCAT. J. Med. Educ., 55: 399-404, 1980.

TABLE 1: MCAT Incremental Validities

Outcome Measure	Statistic	Class Entering: 1976 (Old MCAT)	1977 (Old MCAT)	1978 (New MCAT)
First Year Exam: Total Score	R^2 Non-MCAT	.458	.432	.303
	R^2 added by MCAT	.074	.049	.088
	Total R^2	.532	.481	.391
	Inc. Validity (1)	.16	.11	.29
	Inc. Validity (2)	.14	.09	.13
First Year Exam: National Board Sub-tests	R^2 Non-MCAT	.450	.412	.274
	R^2 added by MCAT	.094	.059	.120
	Total R^2	.544	.471	.394
	Inc. Validity (1)	.21	.14	.44
	Inc. Validity (2)	.17	.10	.17

(Table 1 continues on next page)

TABLE 1: (continued)

Outcome Measure	Statistic	Class Entering: 1976 (Old MCAT)	1977 (Old MCAT)	1978 (New MCAT)
Second Year Exam: Total Score	R^2 Non-MCAT	.443	.342	.334
	R^2 added by MCAT	.072	.041	.095
	Total R^2	.515	.383	.429
	Inc. Validity (1)	.16	.12	.28
	Inc. Validity (2)	.13	.06	.14
Second Year Exam: National Board Sub-tests	R^2 Non-MCAT	.392	.331	.303
	R^2 added by MCAT	.070	.054	.098
	Total R^2	.462	.385	.401
	Inc. Validity (1)	.18	.16	.32
	Inc. Validity (2)	.12	.08	.14

TABLE 2: Incremental Variance Explained by MCAT Component Tests*

Outcome Measure	Class Entering: 1976 (Old MCAT)		1977 (Old MCAT)		1978 (New MCAT)	
First Year Exam: Total Score	Science	(.069)	Science	(.031)	Science Problems	(.061)
	All other	(.005)	GI	(.017)	Biology	(.012)
			All other	(.001)	All other	(.015)
First Year Exam: National Board Sub-tests	Science	(.084)	Science	(.029)	Biology	(.091)
	All other	(.010)	GI	(.015)	Quantitative	(.021)
			All other	(.015)	All other	(.008)
Second Year Exam: Total Score	Science	(.070)	GI	(.029)	Biology	(.054)
	All other	(.002)	Verbal	(.008)	Chemistry	(.024)
			All other	(.004)	All other	(.017)
Second Year Exam: National Board Sub-tests	Science	(.068)	GI	(.030)	Biology	(.059)
	All other	(.002)	Quantitative	(.015)	Chemistry	(.027)
			All other	(.009)	All other	(.012)

*Cells give amount of additional variance explained by the MCAT component tests entered in stepwise regression procedure after all non-MCAT variables.

The Relationship Between MCAT Science Subtest Scores and Performance In Medical School - The Impact of the Undergraduate Institution.

Carter Zeleznik, Ph. D., Jon Veloski, M.A., Samuel Conly, Jr., M.D., and Mohammadreza Hojat, M.A., Jefferson Medical College.

Introduction. This study is concerned with examining the predictive validity of the Medical College Admissions Test (MCAT) science subtest in relation to measures of student performance in medical school and taking into account the undergraduate college which students had earlier attended. Many studies have been reported concerned with the predictive validity of the MCAT (1, 2, 3, 4, 5, 6, 7, 8, 9). Results have been mixed but, in general, low correlations have been found between MCAT scores and performance in medical school during the first two years and very low correlations with performance during the second two years. This has led to a number of criticisms of the MCAT and suggestions about limitations as to its appropriate usage (10, 11, 12, 13, 14). Similarly, there have been justifications of the MCAT (5, 15). In fact, it may be argued that the New MCAT was designed partly in response to the criticisms. If, however, it can be demonstrated that the predictive validity of the MCAT is affected by such factors as the undergraduate institution which the student has attended, it may be appropriate to reexamine some of the notions about the test and what it measures and how it should be used.

It is recognized that institutions have used the MCAT in different ways and that it may be difficult to generalize. One of the main purposes of the MCAT was to discriminate future performance of applicants who might be considered marginal rather than to identify individuals likely to perform at the top level. It is often asserted that the purpose of the MCAT was only to demonstrate applicants' possession of sufficient competence in various areas to permit them to complete the first two years of medical school. Nevertheless, it appears that medical schools have utilized MCAT data in such a way that individuals who have higher scores are more likely to be accepted for admission than individuals with lower scores (16, 17, 18, 19, 20). The validity of these practices, of course, depends upon assumptions concerning the predictive validity of the selection instrument, particular the science subtest of the MCAT insofar as it has been heavily weighted. Moreover, it is common practice for students' grade point averages in college to be weighted in relation to college in order to accommodate differences in grading standards between institutions (3, 21, 22, 23). In arguing for separate MCAT test scores for each of the science areas covered by the test, however, Funkenstein in 196[illegible] suggested that admissions committees might make allowance for students' test results based upon information that in a particular college a certain science was not well taught (16). Similarly, it has been recognized for some time that there might be changes in MCAT for students who elect to repeat the MCAT, especially if some required science course work is completed between tests taken (24). Certainly,

Send reprint requests to: Carter Zeleznik, Ph. D., Associate Director, Office of Medical Education, Jefferson Medical College, 1025 Walnut St., Philadelphia, Pennsylvania, 19107

if there are subgroup differences within the population of medical students which affect the predictive validity of the selection instrument, these should be recognized and taken into account.

Methods. The Jefferson Medical College has established a Cooperative Longitudinal Study, funded by the Josiah Macy, Jr., Foundation, in association with eight undergraduate colleges and universities from which it receives a large number of its applicants and admits a large number of its students. These institutions include both state and private colleges and universities and one Ivy League school. As part of the overall program of sharing data with these institutions, summary statistics were computed concerning mean MCAT scores, National Board examination scores, and other measures. Data were collected for the period covered by entering students between 1964 and 1977. During this period the "old" MCAT was in use. Table 1 provides a summary of means and standard deviations for each institution for science MCAT scores and for part I and part II of the National Board examinations. Separate computations were made for students participating in a five-year accelerated program conducted in association with one of the institutions. Schools are not identified and are grouped in terms of numbers of students included in the calculations. Visual examination of the histograms for the MCAT scores of each institution suggested that they were distributed normally and that restriction in the range of MCAT scores would not interfere with the calculation of meaningful coefficients of correlation. This is also demonstrated by the fact that the standard deviations vary little among colleges. The attrition rate for students during this period has been very low. It is quite likely that utilization of the MCAT for selection purposes contributed to this being the case.

Table I · Means and Standard Deviations
MCAT Science Subtest and Comprehensive Examinations

College	MCAT Science	National Board Examinations I	National Board Examinations II
A	631 (67)	552 (81)	572 (79)
B	643 (68)	530 (94)	567 (92)
C	607 (65)	535 (84)	547 (90)
D	647 (73)	533 (73)	534 (84)
E	642 (61)	523 (84)	515 (88)
F	616 (60)	520 (78)	523 (101)
G	597 (66)	509 (101)	534 (92)

H	621 (74)	528 (77)	527 (89)

(Note: Standard deviations are given in parentheses. Colleges A, B, and C had between 80 and 275 graduates matriculate at Jefferson during the period of the study. Colleges D, E, and F had between 41 and 60 such graduates. Colleges G,H, and I had between 20 and 40 such graduates.)

Results. Table II presents the correlations between scores on the MCAT science subtest and freshman grade point average at this institution as well as Parts I and II total scores on the National Board Examination for students by the undergraduate college which they attended.

Table II Product-Moment Correlations For Scores on the MCAT Science Subtest and Selected Scores in Medical School

Colleges	Freshman G.P.A.	NBME Part I Total	NBME Part II Total
A	.24	.42	.46
B	.37	.40	.37
C	.29	.37	.29
D	-.03	.12	.25
E	.55	.41	.51
F	.27	.40	.18
G	.05	.08	.06
H	.25	.12	.16
I	.23	.55	.52
All Colleges	.29	.36	.37

Rank order correlations between pairs of sets of correlations were also computed in order to obtain a measure of consistency in the overall patterns. The rank order correlation between freshman G.P.A. correlation by college and NBME Part I total scores correlation by college was .30. Rank order correlation between freshman G.P.A. and NBME Part II was .32. However, the rank order correlation between NBME Part I and NBME Part II was a surprizing .93.

Discussion. Data obtained in this study indicate that there is considerable variation among schools in the correlations between MCAT science scores and performance measures during medical school. All correlations, except one, were positive but striking differences existed among these correlations. There was considerable consistency in performance in medical school based on MCAT science scores and the particular undergraduate institution the individual attended. This suggests that it is possible to make differential predictions of student success on some measures in medical school on the basis of MCAT science scores and the individual's particular undergraduate institution, this in spite of restriction in range and ceiling effects. Of course, admissions committees would want other data in addition to this, but given the precarious state of the art of predicting student performance in medical school in general, such findings could be of considerable assistance to members of the admissions committee and to academic advisors within the medical school.

Questions arise, of course, as to why such differences might exist among institutions. It is expected that during the course of the Cooperative Longitudinal Study, these matters will be examined further. Attention will be given, for example, to the interaction of S.A.T., MCAT, and undergraduate GPA by institution. Attention will also be directed to examination of data concerning the New MCAT and differential prediction by institution. At the present time, inadequate data were available to make such an analysis meaningful. However, when the total scores for the old MCATs and for new MCATs were summed algebraically for 45 colleges from which Jefferson receives students and which prepare at least 25 students per year for medical school were correlated by school, the resulting correlation was .83, suggesting at least a gross similarity in relative success on MCATs according to school as has been shown in the past. (These tabulations were performed by Hyman Menduke, Ph. D., of Jefferson Medical College, using AMCAS data.)

Various explanations for the present findings may be considered. It is known that at some institutions advisors encourage students to repeat the MCAT if they do not do well initially whereas at other institutions this is not done. It is not unlikely that at some schools there is greater use made of review courses. Differences in how various courses are taught within and between schools has been mentioned above and could be a factor. It is possible that at some institutions, an implicit goal of instruction is to prepare students for the MCAT in order to enhance the likelihood of their being admitted to medical school without regard to meeting their needs in terms of their later medical school progress.* The time when students take certain science courses in some colleges may be a factor as well. It is also possible that there may be changes observed over time at different institutions. In order to answer many of the questions raised by this study, it would be desirable to have such data available from a variety of medical schools for individual undergraduate institutions. To the best of our knowledge, however, data of this kind have not been formally shared between medical schools and undergraduate institutions previously.

Conclusion. A study has been described in which data concerning performance on the MCAT and in medical school have been compared in relation to the institution from which the students received their undergraduate degrees. Variation in correlations according to institution suggests the possibility of previously unrecognized factors being involved which may be of interest to admissions personnel and to curriculum planners at undergraduate institutions and premedical advisors. The data suggest that rather than weighting undergraduate GPAs on the basis of the observed mean MCAT scores of students at various institutions as is currently the practice

* This in turn may be related to lack of awareness of the (changing) medical school curriculum.

at several institutions, both undergraduate GPAs and MCAT scores of students at those institutions should be considered in light of the performance of previous students at those institutions at particular medical schools. Such variation may be attributed to a variety of causes. Tests such as the MCAT might be thought of, therefore, not as "nationally standarized" tests but as requiring standardization in relation to many factors including the undergraduate institution students have attended. This approach may therefore also be useful in other professional areas such as pharmacy (25, 26) and other related fields.

References.

1. Roessler, R., M.D. Cognitive And Personality Variables In The Prediction Of Preclinical Performance. Rs (Multiple Correlations) p. 127, RIME, 1978.

2. Edwards, C.N. Faculty Assessment Of Medical School Criteria p. 137, RIME, 1978.

3. Friedman, C.P. A Non-Linear Discriminant Analytic Approach to Predicting Medical Student Performance. p.146,150, RIME, 1979.

4. Woods, B., Jacobsen, M., Netsky, M. Social Class and Academic Performance by Medical Students. JME, 42, 225-230, 1967.

5. Sanazaro, P. and Hutchins, E. The Orgin and Rationale of the Medical College Admission Test. JME, 38, 1044-1050, 1963.

6. Garfield, S. and Wolpin, M. MCAT Scores and Continuation in Medical School. JME, 36, 888-891, 1961.

7. Johnson, D., Hutchins, E. Doctor or Dropout? A Study of Medical Student Attrition. JME, 41, 1097-1296, 1966.

8. Erdman, J., Mattson, D., Hutton, J., and Wallace, W. The Medical College Admission Test: Past, Present, Future. JME, 46, 937-946, 1971.

9. Hoffman, E.L., Wing, C.W. and Leiff, H. Short and Long-Term Predictions About Medical Students. JME, 38, 852-857, 1963.

10. Howell, M., Vincent, J. The Medical College Sdmission Test as Related to Achievement Tests in Medicine and to Supervisory Evaluation of Clinical Physicians. JME, 42, 1037-1044, 1967.

11. Bartlett, J. Medical School and Career Performances of Medical Students with Low Medical College Admissions Test Scores. JME, 42, 231-237, 1967.

12. Murden, R., Galloway, G., Reid, J., and Colwill, J. Academic and Personal Predictors of Clinical Success in Medical School. JME, 53, 711-719, 1978.

13. Molidor, J.B. A Factor Analytic Study Of The Old And New MCAT Examinations. p. 141, RIME, 1979.

14. Gough, H., Hall, W., and Harris, R. Admissions Procedures as Forecasters of Performance in Medical Training, JME, 38, 983-998, 1963.

15. New Medical College Admission Test Interpretive Manual. Association of American Medical Colleges, 1977.

16. Funkenstein, D. Testing the Scientific Achievement and Ability of Applicants to Medical School. The Problem and a Proposal. JME, 41, 120-134, 1966.

17. Stefanu, C., and Farmer, A., Jr. The Differential Predictive Validity of Science MCAT in the Admissions Process. JME, 46, 461-463, 1971.

18. Williams, J.H., Pengon, R., Sachs, L., Meleca, C.B., and Camiscioni, J. A Computer-Assisted Admissions Process. JME, 52, 384-389, 1977.

19. Hamberg, R., Swauson, A., and Dohner, C. Perceptions and Usage of Predictive Data for Medical School Admissions. JME, 46, 959-963, 1971.

20. Gee, H., Cowles, J. (Eds) The Appraisal of Applicants to Medical Schools. Jme, 32, Part 2, October, 1957.

21. Johnson, D. A Multifactor Method of Evaluating Medical School Applicants. JME, 37, 656-665, 1962.

22. Clapp, T., and Reid, J. Institutional Selectivity as a Predictor of Applicant Selection and Success in Medical School. JME, 51, 850-852, 1976.

23. Best, W.R., Diekma, A.J., Fisher, L., Smith, N. Multivariate Predictors in Selecting Medical Students. JME, 46, 43-50, 1971.

24. Morris, W.W. The MCAT-Science Survey: Interpreting the MCAT Subtest Scores of Repeaters. JME, 42, 918-925, 1967.

25. Munson, J.W., Bourne, D.W. Pharmacy College Admission Test (PCAT) as a Predictor of academis success. II: A progress report. Am. J. Pharm. Educ., 41(3): 272-4, Aug., 1977.

26. Lowenthal, W., Wersin, J., Smith, H.L. Predictors of success in pharmacy school: PCAT vs. Other admission criteria. Am. J. Pharm. Educ. 41(3): 267-9, Aug., 1977.

Restriction of Range and the Predictive Validity of the New Medical College Admission Test

Thomas J. Cullen, Ph.D., Charles W. Dohner, Ph.D., Percy D. Peckham, Ph.D., Werner E. Samson, Ph.D., University of Washington School of Medicine, Seattle, Washington 98195.

INTRODUCTION:

The issue of predictive validity of the New Medical College Admissions Test (New MCAT) is of major importance now that three classes have presented scores to admissions committees for consideration. In spite of the extensive testing of the new test most committees were, up to this point, at a loss as to the interpretation and use of the scores.

Validity studies have begun to appear (1,2,3) and these early results are encouraging. Nonetheless, as admissions committees use the New MCAT as one variable in the admissions process, it is inevitable that restriction of range of talent will occur on that variable. This means that it is likely that the tests scores for the accepted group at any given school will be concentrated in a small portion of the possible range of scores (4). This phenomomenon will serve to attenuate the validity coefficients calculated for the New MCAT and the selected criterion.

The purpose of this study was to investigate the effect of restriction of range on the validity coefficients calculated for the New MCAT subtest scores and the First Year Comprehensive Examination at the University of Washington School of Medicine. (5) This examination has reliability indices ranging from .89 to .93 (Kuder-Richardson 20). The total score of the examination has correlated with the total score of the National Boards Part I examination in the range of .70 to .75. (6)

The second purpose was to investigate the contribution of each subtest of the New MCAT to the prediction of the total score on the First Year Comprehensive Examination.

METHOD:

Pearson correlations between scores on the New Medical College Admissions Test and the total score on the First Year Comprehensive Examination were calculated for 170 students in the Entering 1979 class. The resultant coefficients were then corrected for attenuation attributed to restriction of range using the following formula.

$$R_{12} = \frac{r_{12}\frac{S_1}{s_1}}{\sqrt{1 - r_{12}^2 + r_{12}^2\frac{S_1^2}{s_1^2}}}$$

Requests for reprints should be addressed to Thomas J. Cullen, Ph.D., E312 Health Sciences Bldg. SC-45, Office of Research in Medical Education, University of Washington School of Medicine, Seattle, WA 98195.

$\frac{S_1}{s_1}$ is the ratio of the standard deviation of the nonrestricted group (including those selected) to the standard deviation of the restricted group on the predictor test; r_{12} is the actual obtained validity coefficient for the restricted group; R_{12} is the estimated validity coefficient for the total, nonrestricted sample. S_1 in this study equals 2. The mean score for the Comprehensive exam was 280.9 with an SD of 26.

A stepwise multiple regression was performed using the six subscores for the New MCAT, Chemistry (C); Biology (B); Physics (P); Science Problems (SP); Skills Analysis:Reading (R) and Skills Analysis:Quantitative (Q) as independent variables and the total score on the First Year Comprehensive Examination (TOT) as the dependent variable.

RESULTS:

Table I illustrates the correlation coefficients for the New MCAT subscores and the First Year Comprehensive Examination total score. As can be seen a substantial increase in predictive validity is achieved by applying the correction formula.

Table II presents the results of the regression analysis. A multiple R of .51 is achieved after all variables have been entered.

DISCUSSION:

It just happens to be the current state of the test constructor's art that validity coefficients as high as .60 are rarely achieved (7). Given this consideration, the corrected coefficients; .63 for C; .60 for SP; and .50 for B are encouraging results indeed. The multiple R of .51 achieved for the regression study is also significant. It indicates that performance on the First Year Comprehensive Exam can be substantially predicted using the New MCATs. Sedlacek (8) has shown that the regression coefficient can be affected by restriction of range in the criterion. If the multiple R were corrected, it would be even higher.

CONCLUSION:

The data demonstrate simple correclations between the New MCATs and the criterion variable -- total score on the First Year Comprehensive Examination-- are substantially enhanced when corrected for restriction of range on the criterion. When investigating the predictive validity of the New MCAT test and, for that matter, any other test, care should be taken to consider the effects of restriction of range and other circumstances that tend to attenuate correlation coefficients. The New MCAT appears to measure well what it is supposed to measure, at least for this sample of students.

TABLE I

Correlations Between the New MCAT Subtest Scores and the Total Score on the First Year Comprehensive Examinations

	N	SD	Pearson r	r Corrected
Biology	170	1.5	.35	.50
Chemistry	170	1.6	.47	.63
Physics	170	2.0	.31	.37
Science Problems	170	1.6	.45	.60
Skills Analysis:Reading	170	1.6	.24	.32
Skills Analysis:Quantitative	170	1.7	.28	.35

TABLE II

Summary of the Regression Analysis

(Dependent Variable = Total Score on the First Year Comprehensive Exam, Independent Variable = Scores on subtests of the New MCAT)

Variables Entered	Multiple r	R^2	Simple r	F	Significance
Chemistry	.4683	.2193	.4683	47.1936	.0001
Biology	.4908	.2409	.3478	26.4950	.0001
Skills Analysis:Reading	.5063	.2564	.2393	19.0768	.0001
Science Problems	.5112	.2613	.4472	14.5931	.0001
Skills Analysis:Quantitative	.5120	.2622	.2805	11.6551	.0001
Physics	.5122	.2624	.3119	9.6625	.0001

Regression Equation: $Tot' = 171.79 + 4.54C + 2.01B + 1.82SA{:}R + 2.07SP + .61SA{:}Q - .28P$

Bibliography

1. Cullen, T.J., et al. "Predicting First-Quarter Test Scores from the New Medical College Admission Test." JME, 55:393-398, 1980.

2. Friedman, C.P., Bakewell, W.E., Jr. "Incremental Validity of the New MCAT." JME, 55:399-404, 1980.

3. McGuire, F.L. "The New MCAT and Medical Student Performance." JME, 55:405-408, 1980.

4. Kaufman, A.S. "Restriction of Range: Questions and Answers." Test Services Bulletin #59, May 1972. The Psychological Corporation.

5. Cullen, T.J., et al. "The Development and Use of A Comprehensive Test For Evaluating Decentralized Medical Education--the WAMI Experience." RIME Conference Proceedings, 1977.

6. Cullen, T.J. Unpublished manuscript "Predicting National Board Part I Performance."

7. Seder, R.J. "The Analysis of Data. Scott, Foresman and Co., 1969, Page 434.

8. Sedlacek, W.E., Hutchins, E.B. "An Empirical Demonstration of Restriction of Range of Artifacts in Validity Studies of the Medical College Admission Test." JME, 41:222-228, 1966.

PRECIS

DEVELOPMENT OF PATIENT ATTITUDES

MODERATOR: Miriam S. Willey, Ph.D.
Howard University, College of Medicine

MEDICAL STUDENTS' ATTITUDES TOWARDS PATIENT'S PHYSICAL, PSYCHOLOGICAL AND HEALTH STATE CHARACTERISTICS

The study investigated medical students expressed willingness to treat patients exhibiting certain physical, psychological or health state characteristics. Highly negative and positive characteristics were identified and the influence of the educational process was investigated.

DEVELOPING A PSYCHOSOCIAL EDUCATIONAL PROGRAM FOR PRIMARY CARE PHYSICIANS: NEEDS ASSESSMENT AND EVALUATION BASELINE

Using screening instruments and interviews with 150 adult medical outpatients at the University of Alabama in Birmingham, the investigators document a wide variety of psychosocial and psychiatric educational needs of primary care internal medicine housestaff. A psychiatric liaison educational program was developed from the needs uncovered by this study and future evaluation of the program will use this data as a baseline.

FOSTERING EMOTIONAL DEFENSIVENESS IN INTENSIVE CARE UNIT RESIDENTS

In a study of a large pediatric intensive care unit it was found that the language employed by the ICU team unnecessarily emphasizes uncertainty and that parents are not routinely recognized as important participants in decisions about medical care of their children. We believe these defensive strategies employed by ICU physicians may not serve the best interests of doctors in training or their patients.

MEDICAL STUDENTS' ATTITUDES TOWARDS PATIENT'S PHYSICAL, PSYCHOLOGICAL AND HEALTH STATE CHARACTERISTICS

C. Warner Johnson, M.D., Kaaren I. Hoffman, Ph.D.
University of Southern California

PURPOSE AND BACKGROUND:

Medical educators have shown a special and ongoing interest in the attitudinal development of their students. This interest is well deserved because attitudes and expectations are believed to strongly influence important student career decisions such as specialty choice, practice location, type of patient relationship and strategies of health care. The present study hoped to provide more specific information on medical students' attitudes by seeking their responses to various patient characteristics.

Following Eron's (1) lead, a large proportion of the previous research deals with the general attitudes, such as cynicism, and social concerns of medical students. Rezler (2) in a review of this area concludes that medical school probably does contribute to increased cynicism and at best has no effect on such attitudes as humanism or benevolence.

Another large area of investigation reviewed by Rezler concerns the effect of various teaching programs on the student's attitudes toward comprehensive care (i.e., care which encompasses a patient's emotional and family problems as well as his organic pathology). Rezler concludes that any positive orientation toward treating patients with social and emotional problems is difficult to obtain once students have developed a preference for the physically ill and is most likely short-lived once students have returned to the regular medical school environment.

Recent research is of a more specific nature, tending to deal with attitudes as they affect the major health problems of the day. Studies have been conducted on the attitudes toward certain disease states such as cancer (3) and the chronically ill (4) and on the attitudes believed to have a direct affect on the delivery and quality of health care found in society as a whole (5,6).

The present paper seeks to relate attitudes to health care delivery by measuring students willingness to treat patients exhibiting various characteristics. Previous data (7,9,10) have indicated that students have definite attitudinal biases toward certain patient types. Here, however, we are interested in examining the individual physical, psychological and health state characteristics of patients which may or may not combine to form a patient type. In this regard we concur with Leserman (5) who suggested that the more specific factors which one can identify the more likely the correlation between attitudes and behavior and the more likely corrective actions can be taken. In particular, we are interested in identifying which characteristics evoke highly negative or positive reactions, whether such reactions undergo change during the education process, and if so, when these changes occur.

Address reprint requests to: C. Warner Johnson, Director Medical Student Education in Psychiatry, University of Southern California School of Medicine, 2025 Zonal Avenue, Hoffman 101, Los Angeles, California 90033.

The authors wish to acknowledge the valuable contribution of Frank Donnelly, Ph.D. in the formulation and development of this research.

METHODOLOGY AND INSTRUMENTATION: The data gathering instrument consisted of a list of 67 patient characteristics. The list is wide ranging and includes physical, social, emotional and personality characteristics. The list was generated from a survey of third and fourth year students which asked them to list separately both the three characteristics they most and least liked their patients to have. The data instrument was pretested on several groups of students rotating through the psychiatric clerkship and some modifications were made to eliminate any ambiguities.

For each of the 67 characteristics, students were asked to indicate their willingness to treat, in terms of their practice time, patients manifesting the characteristics. They indicated whether they would be willing to (1) mostly treat, (2) regularly treat, (3) occasionally treat, or (4) rarely treat such patients. The summation of the data, then, represents the percent of medical students who would respond to each given patient characteristic in the four categories above. For purposes of analysis the data was collapsed into two categories by combining the mostly and regularly treat categories and comparing them to the combination of the occasional and rarely treat categories.

The data instrument was administered at the beginning of the year to all Year I and Year II students at the USC School of Medicine. Data for the third and fourth year students was gathered throughout the year at the beginning of their psychiatric clerkship. This same procedure also was used to gather data from students participating in a comparable psychiatric clerkship at UC Irvine Medical School.

It is understood that when one performs statistical tests on each of 67 characteristics, some of the results will be significant by chance alone. Therefore, to validate any noted changes in students' response across the years as well as to pinpoint precisely when (at what point in the educational process) the changes occurred, a short form of the questionnaire was developed and re-administered to various groups. The short form consisted of 35 patient characteristics, and included those which had shown a significant change across the years. Eight characteristics which showed no change in the primary analysis were also randomly included.

The short form was administered to the Year II students at the end of their second year and to the Year III students at the end of their third year of medical school. One final administration of the short form involved retesting a select group of students in the second year as they started their third year. This final sampling took place at the beginning of the first two psychiatric clerkships of the third training year, approximately two and eight weeks, respectively, after the administration conducted at the end of the second year mentioned above.

The overall data design involves comparing data from different class groups gathered at the same time (a cross-sectional approach) as well as comparing data from the same group gathered at different times (a longitudinal approach). Table 1 summarizes the various administrations of the long and short data forms and indicates which class groups were involved.

RESULTS: Data was gathered as to the percent of students at each given year level who expressed a willingness to regularly treat patients exhibiting each of the 67 characteristics.* Chi-square analyses between the two schools showed

* Complete information on the students' responses for each year level and administration time will be provided in handouts at the time of presentation.

them to be quite comparable. No significant differences were noted between the Year II students at the two schools. For Year III students, three of the 67 characteristics, "marital-family problems", "non-English speaking" and "poor historian", showed a difference significant at the .05 level.

Those characteristics which evoked a substantial negative response by the time of the third year of medical school (less than 40% of students willing to regularly treat) are listed in Table 2. For 12 of these 26 characteristics (denoted by an asterik) the negative reaction seemed to be present at the time of entrance to medical school. Only one characteristic, "requires pain indu-therapy", which evoked a strong negative reaction at the start of medical school improved significantly across the years.

The vast majority of third year students (90% or more) expressed a willingness to treat patients exhibiting the characteristics denoted in Table 3. All of these seven characteristics were positive at the start of medical school Only two of these characteristics, "multiple, current diseases" and "low income", underwent a significant decline.

Table 4 lists characteristics which showed a significant difference in the expressed willingness to treat across the three medical school years. The major difference occurred between Years II and III. Only two characteristics, items 11 and 27 in Table 4, showed a significant difference between Years I and II. Twenty-seven of the 34 listed characteristics showed a substantial drop between the second and third years. Four characteristics, numbers 2, 12, 22, and 25, seem to show significance due to a discrepant value in Year II and three characteristics, numbers 11, 16, and 27, seem to show significant change due to discrepant values in Year I.

The first 28 characteristics noted in Table 4 showed significance in the primary analysis. Confirmation of the changes noted in these items was sought by incorporating them into another data form (short form) along with randomly selected characteristics which had shown no significant difference across the years. This short form of the data was then re-administered to the Year II and III students at the end of the school year. Analysis showed no significant differences in the data gathered at the beginning and end of the second year. Data from the beginning and end of the third year did show significant differences. A number of characteristics (18 in all) continued to show a significant decline in the third year (i.e., fewer students at the end of the year were willing to regularly treat patients exhibiting these characteristics). Confirmation of the differences across years was sought by comparing the data gathered at the end of Year II and the end of Year III. In all cases confirmation was found and several additional items (numbers 29 to 34 in Table 4) now showed a significant difference.

To pinpoint precisely where the major change in willingness to treat occurs, data was gathered from some 38 Year II students as they started their Year III clerkships. Analysis shows that at least 15 characteristics which showed a difference between Years II and III show that difference almost immediately after beginning the third training year ($p<10$).

DISCUSSION: One major purpose of the present paper was to identify characteristics which evoked a highly negative reaction. Such characteristics are listed in Table 2. Although there is considerable overlap, these characteristics can be seen to cluster into four areas. One cluster involves characteristics which hinder effective communication in the doctor-patient relationship. Some of these, e.g., non-English speaking, repetition of symptoms, poor historian, might very well be easy to moderate with education in interview techniques and variation in clinical procedures. Others, such as delusions, bizarre behavior, hallucinations, etc., might be more difficult to alter. These latter characteristics have psychological overtones and might also be a reflection of the student's preference for physical versus psychological problems previously reported on (2,10).

A second cluster of characteristics verifies the suggestion of previous research that patients without a diagnostic problem including cancer and chronically ill patients tend to be avoided (3,4,10). Indications are that such patients offer less opportunity for feelings of "success" or educational gain (9,11). The findings here and elsewhere emphasize the need for educators to respond to student motivations and needs for appreciation and successful treatment outcome. Realistic expectations for self and patient need to be formulated.

A third cluster of characteristics might simply be a reflection of a general societal rejection of certain characteristics. Such characteristics as unappreciative, irritable, angry, hostile, unpleasant odor, and complaining, whining fall within this cluster. These characteristics are distinguished by the fact that they evoke a negative reaction upon entrance to medical school and seem to remain constant throughout. Fortunately, the frequency of encountering such characteristics might very well be minimal.

Societal values might also play a large part in the definition of a final cluster of patient characteristics, those epitomizing self-indulgent or self-destructive qualities. Averse reactions are noted to drug and alcohol dependency, obesity, attempted suicide and low pain tolerance. This cluster is distinguished from that mentioned above, in that the reaction to these characteristics tends to become more negative during medical school. Such progressive aversion occurs even though many patients exhibiting such characteristics present with complicated, interesting medical problems secondary to their drug/alcohol abuse. To the extent that these reactions reflect those of society, they may very well be highly resistant to change and need intensive educational efforts. Such effort, however, might well be essential in that these are characteristics of patients who possess major and commonly encountered disorders which most physicians must treat. Certainly it is well recognized that obesity, alcohol and drug abuse are among the leading health problems in the United States and suicide is among the top ten causes of death.

The second major purpose of the present paper was to assess the influence of the medical education process on students reactions to the patient's characteristics. Here the results clearly show that several characteristics undergo a significant change in acceptability and that this change occurs primarily at the start of or during the clinical years. Particularly interesting is that the changes were all in a negative direction and that many occurred almost immediately, within two or three weeks of the beginning of a clinical rotation. The change in student attitudes toward certain patient characteristics, therefore, must be closely connected with the assumption of clinical responsibility. Several hypotheses can be made. Perhaps, the patients actually encountered do not meet the students expectations. Patient contact prior to the clinical years is often quite selective and not necessarily reflective of the patient population as a whole. It is possible that far more of the patients than expected exhibit characteristics already perceived as negative by the students. This lack of concordance with expectations might produce a more generalized negative reaction.

A second possible explanation involves the phenomenon of "modeling" or the taking on of the attitudes, values, and behavior of the group to which one belongs. This implies that some of the negative attitudes might actually be indirectly learned from other staff members. The curriculum itself might also serve to foster some negative attitudes. Limited time on any given rotational medical service precludes observing or experiencing success with patients whose response to treatment is slow or not immediately apparent. This lack of a "success" experience is believed of particular importance with the psychiatric or chronically ill patient (4).

One final possible explanation for the noted attitudinal change centers on the educational process. Several studies have documented the fact that the patient is or can be viewed as an educational object (8,11). The clinical years involve tremendous time and work pressures to "produce" and to please instructors. In this atmosphere, it is likely that anything which interferes with the diagnostic/treatment process such as inadequate data base, poor historian, etc., would escalate the student's anxiety and hence aggravation.

Future research is obviously needed to attempt to examine the role of these various hypotheses. It is highly probable that all may operate depending upon the particular patient characteristic examined. Future research is also needed to ascertain the permanence of the observed attitudes and attitudinal change. A previous study (12) indicated that the observed increase in cynicism is partly situational in nature and subsides as physicians enter "high-interaction" specialties. Reaction to patient characteristics may or may not be of a transitory nature.

References:

1. Eron, L.D. The Effect of Medical Education on Attitudes: A Follow-Up Study. J. Med. Educ., 33:25-33, 1958.
2. Rezler, Agnes G. Attitude Changes During Medical School. J. Med. Educ., 49: 1023-1030, 1974.
3. Haley, H.B., Huynh, H., Paiva, R. and Juan I.R. Students Attitudes Toward Cancer: Changes in Medical School. J. Med. Educ., 52: 500-507, 1977.
4. Kutner, Nancy. Medical Students Orientation Toward the Chronically Ill. J. Med. Educ., 53: 111-118, 1978.
5. Leserman, Jane. The Professional Values and Expectations of Medical Students. J. Med. Educ., 53: 330-336, 1978.
6. Liccione, W.J. and McAllister, S. Attitudes of First Year Medical Students Toward Rural Medical Practice. J. Med. Educ., 49: 449-451, 1974.
7. Reynolds, R.E. and Bice, T.W. Attitudes of Medical Interns Toward Patients and Health Professionals. J. Health Soc. Behav., 12: 307-311, 1971.
8. Merton, R., Reader, G.G., and Kendall, P.L. The Student Physician. Cambridge, Massachusetts: Harvard University Press, 1957.
9. Becker, H.S., Geer, B., Hughes, E.C., and Strauss, A.C. Boys in White. Chicago: University of Chicago Press, 1961.
10. Stoller, R.J. and Geertsma, R.H. Measurement of Medical Students' Acceptance of Emotionally Ill Patients. J. Med. Educ., 33: 585-590, 1958.
11. McGlynn, T.J., Wynn, J.B. and Muzenrider, R.F. Resident Education in Primary Care: How Residents Learn. J. Med. Educ., 53: 973-981, 1978.
12. Reinhardt, A.M., and Gray, R.M. A Social Psychological Study of Attitude Change in Physicans. J. Med. Educ., 47: 112-117, 1972.

Table 1. Summary of the number of respondents and administration times for Data-forms

	Start 1976-77	During 1976-77	End 1976-77	Start 1977-78
USC Year I	N=136			
USC Year II	N=124		N=82*	N=38*
USC Year III		N=144	N=88*	
Irvine Years II & III		N=40		

* Administration involved short-form questionnaire

Table 2. Characteristics evoking a negative reaction in Third year medical students

* Uncooperative
  Low pain tolerance
  Infant
* ~~Reticent to~~ relate symptoms
  Drug overdose
* Unappreciative
  Unresponse to treatment
  Memory impairment
  Non-English speaking
* Indirectly self-destructive
  Poor historian
  Habituation to alcohol
* Requires pain-inducing therapy
* Intoxicated with alcohol
* Irritable, angry, hostile
* Dellusional
  Passive, dependent
  Disheveled, unkempt
  Terminal illness
* Bizarre behavior
* Unpleasant odor
* Complaining, whiny
  Markedly overweight
  Attempted suicide
  Repeated recitation of symptoms
  Organic brain disease
* Hallucinating
  Habituated to drugs

* Indicates a characteristic which evoked a negative reaction in Year I.

Table 3. Characteristics evoking a positive reaction in Third year medical students.

Middle-aged
Compliant, follows directions
Opposite sex
Same sex
Acute curable illness
Middle income
Well groomed

Table 4. Patient characteristics which undergo a significant change in acceptance during medical school.

1. Multiple, current disease
2. Marital-family problems
3. Adolescence
4. Sexual disturbance.
5. Low pain tolerance
6. Children
7. Infant
8. Drug overdose
9. Poor historian
10. Habituation to alcohol
11. Requires pain inducing therapy
12. Tearful
13. Passive, dependent
14. Disheveled, unkempt
15. High intelligence
16. Skin lesions
17. Bizarre behavior
18. Attempted suicide
19. Repeated recitation of symptoms
20. Markedly underweight
21. Organic brain disease
22. Hallucinating
23. Needful of firm limits - direction
24. Low income
25. Genital disease
26. Pregnancy
27. Requires surgical intervention
28. Non-ambulatory
29. Talkative
30. Unresponse to treatment
31. Non-English speaking
32. Delusional
33. Chronic, controllable illness
34. Depression, sadness

Developing a Psychosocial Educational Program for Primary Care Physicians: Needs Assessment and Evaluation Baseline

Steven A. Cohen-Cole, M.D., M.A., University of Alabama in Birmingham, John Boker, Ph.D., University of Alabama in Birmingham, Julian Bird, M.R.C.P., M.R.C. Psych., University of Alabama in Birmingham, Arthur Freeman, M.D., University of Alabama in Birmingham

Introduction

A wealth of epidemiological, clinical, and experimental data all point to the importance of psychosocial variables in influencing physical health outcomes.[1-3] Recognition of this dimension of medicine, along with the rapid growth of primary care training programs, has led to an increased interest in teaching the evaluation and management of psychosocial aspects of patient care. The research reported in this study was designed as part of an ongoing liaison psychiatry education project in Primary Care Internal Medicine at the University of Alabama in Birmingham. In the interest of developing a suitable curriculum, the authors sought to uncover basic areas of need concerning psychosocial skills in medical practice, as well as establishing baseline competencies for the future evaluation of teaching outcomes.

Review of Literature

Numerous reports point to significant problems in the psychosocial training of physicians. Houpt[4] reviews the extensive literature documenting significant emotional disorders in 20-80% of all medical patients, and Marks[5] and Knights[6] demonstrate that about 30% of the emotional and cognitive disorders in medical patients are missed by primary physicians. Furthermore, reports of significant communication problems between doctors and patients[7-9] may account for part of the problem of patient noncompliance with physician recommendations[10] and patient dissatisfaction with medical care.[11] While many liaison psychiatry programs have been described, all of which focus on psychosocial training of non-psychiatric physicians,[12] our review of the literature revealed just thirteen outcome studies, only two of which assessed patient outcomes[13,14] and none of which systematically evaluated changes in routine physician behaviors.

The present research documents basic areas of need in psychosocial dimensions of patient care, in order to design an appropriate educational curriculum for medical housestaff and to establish a baseline for future evaluation. The data include a survey of emotional disorders, stressors, and supports, communication variables, and variables relating to compliance and satisfaction. The authors maintain that this needs assessment and psychosocial baseline data facilitate the design of an appropriate intervention program, as well as providing multiple physician and patient variables for subsequent outcome evaluation.

Reprint requests to Dr. Cohen-Cole, Assistant Professor of Psychiatry, University of Alabama in Birmingham, Birmingham, Alabama, 35294. The authors wish to thank Drs. Jack Hain, Robert Markush and Alwyn Shugerman for their helpful comments and William Bond and Mike Shehi for their research assistance.

Methods

Subjects. The subjects were 150 medical outpatients studied during the course of their regular visits to the general medicine clinic. Patients included a wide age range of adults (mean age 49), and were predominantly female (71%), black (77%), and of low socioeconomic status (45% earned less than $5000, 87% less than $10,000, and more than 75% were in the lowest three levels of Hollingshead's SES Index[15]). Randomly sampled from among all clinic patients during the three month study, both new (36%) and return (64%) patients were included. About 10% refused to participate.

Materials. Patient-care data were obtained by patients' oral or written responses to a series of questionnaires and oral responses to a structured patient interview. The questionnaires assessed recent life change, social support, emotional disorders, and depression. In addition to these instruments, patient interviews asked about knowledge of diagnoses and recommendations, doctor-patient communication and compliance issues, patient satisfaction, and amount and usefulness of psychosocial data that was discussed.

Recent life changes were measured by a fifteen item abbreviated form of the "Holmes and Rahe Social Readjustment Rating Scale."[3] (Modified by Robert Markush, M.D. at the University of Alabama in Birmingham.) In the interview, patients were asked to estimate the magnitude or severity of the changes upon their life on a 10-point Likert scale.

Social support was evaluated by a 16-item scale developed by the authors. It assessed perceived support as well as satisfaction with overall levels of crisis, family, friendship, and religious support. Items included Likert, bimodal (yes, no), and simple completion questions. A composite social support index was derived for patients by summing their respective z-scores on each item. Pilot studies indicate the social support scale has acceptable item variances, item-total correlations, and internal consistency reliability to warrant further use and study as a general screening instrument.

Probable emotional disorders were assessed by the 28-item General Health Questionnaire (GHQ) developed by Goldberg.[16] The specificity and sensitivity of this scale exceeds 85%: similarly, the reported correlation between GHQ total score and independent psychiatric assessment exceeds +0.75. Probable depression was evaluated by the Center for Epidemiologic Studies' Depression (CESD) scale. The usefulness of the CESD in nonpsychiatric clinical populations is summarized by Comstock and Helsing.[17]

During the interviews, patients' knowledge about diagnoses and recommendations was evaluated by asking patients what their doctors said was the matter with them and what they were told to do about it. Communication and compliance issues were investigated by asking patients if they had any unanswered questions about their illness or treatment, if they felt the doctors told them enough to satisfy them, if they understood what the doctors told them about their problems and what to do about them, and whether they have been following their doctors' recommendations. The amount of psychosocial discussion was assessed by asking patients if their doctors inquired about family life or personal problems they might be experiencing. If they were asked, patients then indicated how much time was spent on psychosocial issues, whether they felt it was helpful, and relatively speaking, whether the time spent was "too much", "too little", or "just about right".

If they were not asked about personal aspects of their lives, patients indicated whether or not they would have wanted their doctors to inquire about psychosocial issues.

Chart reviews also were done for each patient in the study. The primary purpose was to monitor the extent of physician recognition and/or recording of psychosocial variables in patients seen in the general medicine clinic. Aspects evaluated by the chart review included problems and recommendations recorded by the physician, frequency of psychiatric diagnoses and management plans, use of psychotropic drugs, and number of mental status and psychosocial history words.

Procedure. All questionnaire administration, interviews, and chart reviews were done by two trained second-year medical students. Patients were approached in the waiting rooms of the medical clinic area prior to their scheduled appointments. After receiving a description of the project, if they gave written informed consent, they filled out the questionnaires and were subsequently interviewed. Twenty-eight percent of the patients could not self-administer the questionnaires either due to illiteracy or limitations imposed by functional handicaps (eyesight, arthritis, etc). In these instances, the questionnaires were read to the patients who, in turn, responded verbally. Preliminary analyses of patient-care data indicated no systematic differences between the two interviewers on any variable.

Results

Using the recommended cut-offs for the GHQ[16] and the CESD[17] we found that 63% of all medical patients had probable emotional disorders and 48% had probable depression. While total life change was not associated with either GHQ or CESD scores, the number and perceived magnitude of undesirable life events (e.g. deaths in the family, job loss) was significantly correlated with both probable emotional disorder and probable depression (p .001). Low social support was also significantly associated with GHQ and CESD scores ($p < .05$). Furthermore, we found an internal validation of the scales for emotional disorder and depression, in that more than 91% of the patients in the depressed range on the CESD (N = 41) also scored in the disordered range on the GHQ. As would be predicted, the reverse was not true: i.e. a group of patients with probable emotional disorders did not score as probably depressed on the CESD.

From the review of patient charts, it was found that 20% included psychiatric diagnoses, 11% contained psychosocial treatment plans, and 16% of patients were receiving psychotropic drugs (9% antianxiety agents, 5% antidepressants, and 2% antipsychotics). Seventy-eight per cent of patient charts contained no mental status words, and 85% of patient charts contained no psychosocial history. The median number of mental status and psychosocial history words per chart was less than one.

From patient interviews, it was determined that patients, on the average, could report only 56% of the problems listed by their physicians in the medical record. Furthermore, patients could only describe, on the average, 60% of the recommendations they were supposed to be following, according to physicians' notes in the medical records.

Thirty per cent of all patients told the interviewer that they had questions about their problems that they didn't get a chance to ask their doctor. Eighteen per cent of all patients said they did not understand what their doctors told them. Twenty-three per cent of patients said they were not following their doctors' recommendations. Seventy per cent of patients reported that their doctor did not ask about family life or personal problems, and 64% of this group said they would have wanted their doctor to ask about this aspect of their lives.

Sixty-three per cent of the patients who said their doctors did ask about personal or family problems, said that the total time spent was less than ten minutes (it was less than one minute for 21% of these patients). Furthermore, 90% of this group of patients who said they discussed their personal problems with their physician said that the time spent on this subject was "just about right" and 90% also said they found the discussion "helpful." No patient reported spending too much time on this subject.

Discussion

The results of this study corraborate numerous other epidemiological findings of the high prevalence of emotional disorders in general medical populations.[4] The brief life events scale we used yielded an association between negative life events and probable emotional disorders similar to that reported in other studies which used more elaborate instruments.[18,19] Furthermore, our social support scale was also significantly associated with probable emotional disorders and depression. While physical and mental health risks associated with low social support have been previously reported,[20] no generally-usable measurement instrument has yet been developed. The life change and social support instruments developed for this study present the possibility of refinement into educationally and clinically reliable and valid indicators of important health-risk variables in medical patients.

It comes as no surprise to anyone familiar with routine medical care that a substantial proportion of emotional disorders go unrecognized and untreated in medical clinics. Our chart reviews corroborate these findings and indicate that psychological data are not routinely elicited or recorded.

The patient interviews confirm that the absence of psychosocial data in patient charts is not simply due to physicians' falure to record information they may have elicited, since 70% of all patients reported that their doctors did not ask at all about personal or family life.

These failures to recognize, evaluate, and treat emotional disorders in the medical setting are paralleled by major communication problems which affect the quality of patient care, compliance, and most probably, physical health outcomes. Since we found that a large proportion of patients (40%) could not even report their doctor's recommendations, that 30% had questions that weren't answered, and that 20% didn't understand what their doctors told them, it seems clear that the high degree of actual patient failure to adhere to therapeutic advice, (usually estimated at about 50%)[21] may be due as much to physicians' communication deficiencies, as to patient-related variables.

The large majority of patients who said they would have wanted their doctors to ask about personal and family life points to an important unmet need of medical patients. Furthermore, we found that patients are quite content with even very small amounts of perceived time devoted to this subject, when it does occur. Considering that 21% of patients reported less than one minute of psychological exploration, and 63% less than ten minutes, the strikingly high percentage of satisfaction with these discussions (90%) indicates that patients are not insisting on lengthy psychosocial involvement.

Taken as a whole, this study documents several important educational needs for better psychiatric and psychosocial training of internal medicine housestaff. These include the need for increased skills: 1. In the recognition and management of emotional disorders (including depression); 2. In the evaluation and management of patients experiencing negative life events and low social support; 3. In communication with patients about their diagnoses and treatment plans; and 4. In the elicitation and recording of psychosocial history and mental status data.

Our current liaison psychiatry educational program (Bird, J. and Cohen-Cole, S., "New methods in liaison psychiatry education: The behavioral approach," manuscript in preparation) has oriented its didactic, clinical supervision, case conferences, and structured exercises to the explicit implementation of the objectives listed above. The evaluation methodology is itself, in turn, keyed to the measurement of changes in these dimensions of patient care and should provide a useful indicator of the success of the above program.

Conclusion and Implications

The present study reports needs assessment research relating to psychiatric and psychosocial education and evaluation of primary care internal medicine housestaff. By investigating actual patients, through screening instruments and interviews, as well as by reviewing patient medical records, the investigators were able to document several areas of suboptimal medical care and unmet patient needs.

An on-going psychiatric liaison education program for medical housestaff has been organized around these documented needs, and future controlled evaluation studies will include these as a baseline for the assessment of program efficacy.

References

1. Cassel, J., "Psychosocial processes and stress: Theoretical formulation," International Journal of Health Services, 4:471-482, 1974.
2. Engel, George, "The need for a new medical model: A challenge for biomedicine," Science, 196:129-136, 1977.
3. Holmes, T., and Rahe, R., "The social readjustment rating scale," J. Psychosm. Med. 11:213-218, 1967.
4. Houpt, J., Orleans, C., George, L., and Brodie, H., "The role of psychiatric and behavioral factors in the practice of medicine." Am. J. Psychiatry, 137:37-47, 1980.
5. Marks, J., Goldberg, D., and Hillier, V., "Determinants of the ability of general practitioners to detect psychiatric illness," Psychol, Med. 9: 337-353, 1979.
6. Knights, E., and Folstein, M., "Unsuspected emotional and cognitive disturbance in medical patients," Annals Int. Med., 87:723-724, 1977.

7. Platt, F.W., and McMath, J.C., "Clinical hypocompetence: The interview," Anals. Int. Med., 91:898-902, 1979.
8. Golden, J.S., and Johnston, G.D., "Problems of distortion in doctor-patient communications, Psych. in Med., 1:127-149, 1970.
9. Korsch, B.M., Gozzi, E., and Francis, U., "Gaps in doctor-patient communication. I. Doctor-patient interaction and patient satisfaction," Pediatrics, 42:855-871, 1968.
10. Davis, M.S., "Variations in patients' compliance with doctors' advice: An empirical analysis of patterns of communication," A.J.P.H., 58:274-288, 1968.
11. Wolf, M., Putnam, S., James, S., and Stiles, W., "The medical interview satisfaction scale: Development of a scale to measure patient perceptions of physician behavior," J. Beh. Med., 1:391-401, 1978.
12. Greenhill, M., "The development of liaison programs," in Usdin, G., (ed.), Psychiatric Medicine, Brunner/Mazel, 1977, pp. 115-191.
13. Adsett, C., and Rudnick, K., "Psychiatric liaison with family practice teams," Psychiatric Opinion, 15:29-33, 1978.
14. Dubovsky, S., Getto, C., and Paley, J., "Psychiatrists on the coronary care unit," Psychosomatics, 18:18-27, 1977.
15. Hollingshead, A.B., The two-factor index of social position, Privately published, New Haven, Conn., 1957.
16. Goldberg, D., Manual of the General Health Questionnaire, NFER Publishing Company, Windsor, Great Britain, 1978.
17. Comstock, G.W., and Helsing, K.J., "Symptoms of depression in two communities," Psychol. Med., 6:551-563, 1976.
18. Rahe, R., "Life change events and mental illness: An overview," J. of Human Stress, 5:1-10, 1979.
19. Lloyd, C., "Life events and depressive disorder reviewed," Arch. Gen. Psych., 37:529-552, 1980.
20. Cobb, S., "Social support as a moderator of life stress," Psychosom. Med., 38:300-314, 1976.
21. Sackett, D.L., and Haynes, R.B. Compliance with Therapeutic Regimens, Baltimore, Johns Hopkins University Press, 1976.

Fostering Emotional Defensiveness in Intensive Care Unit Residents

by: Joel Frader, M.D., Charles Bosk, Ph.D., and Ellen Prince, Ph.D.
University of Pennsylvania

Purpose

Following a pilot project investigating some difficulties inherent in providing and teaching about pediatric intensive care, we undertook a comprehensive sociologic and linguistic study of a large pediatric intensive care unit (ICU). The preliminary study led us to believe that technology dominates the ICU. We therefore set out to collect detailed information about the activities of and interaction among the physicians working in the unit. The purpose of this effort was to uncover how trainees in the ICU learn attitudes and behaviors which favor narrow technologic, rather than broad humanistic, approaches to critically ill children and their families.

Review of Literature

A comprehensive review of much of the literature relevant to this area has recently been published by one of us.[1] In that review, we pointed out that little substantive information was available. The existing literature only touches on the broad problems of nonbiomedical aspects of working and learning in an ICU.

To that review, must be added mention of two recent papers, as well as a look at the pertinent linguistic literature. Waller and his group[2], reported the emotional conflict experienced by families of critically ill children who had been informed of poor prognoses. They noted that some parents became hostile when their denial was confronted directly. They also reminded us that prophesying doom may itself be a defense employed by physicians. In 1 of 4 cases, a miraculous recovery occurred, thus highlighting the importance of uncertainty in the medical enterprise.

Youngner and colleagues[3], in a study of house staff and nurses in a medical ICU, found that the two groups had remarkably similar views about ethical problems and decision making in the unit. Nonetheless, physicians felt that they, more than patients, family, or the ICU staff as a whole, should make decisions regarding stopping/withholding treatment for patients with life threatening illness.

In the linguistic literature relevant to our study, workers have shown that speakers give formally identifiable cues about whether they believe hearers share knowledge or understand the context of what is being spoken

---

Address for reprints: Joel Frader, M.D., The Children's Hospital of Phila. 34th & Civic Center Blvd., Philadelphia, PA 19104

Partial support for this work came from The Robert Wood Johnson Foundation (Clinical Scholars Program), The National Fund for Medical Education (Fellowship Award), the Health Services Administration (Grant MC-R-420423).

about. That is, the use of certain phrases or words allows speakers and hearers to communicate without continuously identifying and re-defining the subjects(s) of their discourse.

However, certain types of speech may be termed "fuzzy communication." This "hedging" makes it more difficult to appreciate whether all parties to communication interpret the same words in equivalent ways.[15-17] Two researchers in a medical environment have recently reported great variability among pathologists, radiologists, other physicians, and students in understanding commonly employed medical hedges, such as "consistent with" or "normally."[18]

Methods

We employed two approaches to this phase of our investigation: 1) the social science field work technique of participant observation and 2) tape-recording of daily rounds in the ICU with subsequent verbatim transcription and detailed linguistic analysis. Participant observation was carried-out by two of us (JF and CB). Lingusistic analysis was done by the other author (EP).

The observations were made during morning rounds and at times we accompanied house officers throughout the day. These experiences were recorded in field notes. The notes constitute one set of data used in the study.

Rounds in the intensive care unit were taped and transcribed after participants were notified of the project and permitted to decline cooperation with the study. Recording equipment was set-up and tapes made over a six month period. From the technically adequate recordings, tapes of entire sessions were selected for transcription. Thus, participants in rounds did not know which tapes were to be transcribed for analysis. This was done to minimize any observer effect.

Transcripts were made by two experienced graduate students in linguistics. They were reviewed by the pediatrician for accuracy. The identification of speakers was coded to preserve anonymity and minimize bias in analysis. References to specific patients, families, medical personnel, and institutions were also altered to prevent misuse of the data.

Results

First, our observations led us to believe that medical-technical uncertainty was a source of frustration for the ICU team. When team members, especially house staff, found they were unable to accurately predict patient outcomes (survival and/or quality of life), they stopped committing themselves about prognosis. In fact, on close examination of the transcripts, we found that verbal uncertainty is the most pervasive characteristic of all of the recorded discussions. ICU team members hedge (that is, use phrases that reduce clarity and definition) between 150 and 350 times an hour. This degree of hedging is excessive in comparison to that in other "natural language" conversations studied by linguists.

The hedging in the pediatric ICU is of two varieties: there are "approximators" and "shields."

231

Approximators are used to indicate uncertainty about the content of spoken assertions. One hears, for example, that a child's feet are "sort of blue." This kind of hedging is most common in descriptions of signs and symptoms or, oddly enough, in association with numerical expressions. An example of the latter is "the inspired oxygen concentration was about fifty-four percent." In these cases, the speaker is committed to the truth of some degree of his or her description. The uncertainty involves the precision of the statement.

Shields involve the establishment of distance between the speaker and the spoken statement. In the case of the blue feet, one might hear "I think the feet are blue." Here, the speakers invite questions about the plausibility of the statements. These hedges are used throughout the discussions, but are most noticable when an assertion is attributed to another person.

It is interesting to note that both types of hedges are used even when precise technical information is available for clarification or to settle disputes.

Our second major finding was that crucial treatment decisions--even those involving surgical intervention or use of sophisticated life supporting technology--were often discussed without mention of the patients' families. In over 2 1/2 hours of conversation at rounds from one week in the ICU, we found 14 discrete references to parents. On 11 occasions, the families were passively regarded as providers or recipients of information. Examples of these parental references were "...her parents say this is pretty much baseline for her..." or "The mother called, she wanted t'know...."

In 1 case, the parents had asked that a child being discharged from the intermediate care portion of the ICU be sent home on a Sunday, rather than Friday, for "logistical reasons." The parents wishes in this case were respected.

In only 2 cases, both involving the same baby who was expected to die, did the team members hold parental thoughts or feelings to be of major consequence. Even in those parts of the conversation, the parents were assigned a rather passive role: "...we're gonna try to impress on the parents how sick the baby is and possibly talk about making some decisions."

Discussions and Conclusions

We believe that trainees in the pediatric ICU learn to hedge verbal expressions, even when medical circumstances do not require linguistic uncertainty. Members of the ICU team are psychologically unable or reluctant to be definitive. We are concerned that this use of language reflects a desire to avoid responsibility for reporting unsettling facts or taking actions.

House officers also do not learn systematic appraisal and consideration of the role of the family in the care of critically ill children. This reality is different from the belief they have expressed (in another phase of the study) that parents should have a place in the decisions in the ICU.

The unconscious lessons the trainees learn in the pediatric ICU require some careful reflection. On the one hand, we recognize the importance of psychological defense mechanisms in coping with a stressful environment, such as the ICU. However, we question whether this particular education fosters the best possible care for the patients and the families.

It is important to point out that the above ideas have not escaped the attention of those responsible for patient care and house staff education in the pediatric ICU in our hospital. We have embarked on an educational effort to improve pyschosocial and ethical aspects of care in our unit. In the future, we hope to be able to report on the efficacy of that enterprise.

Bibliography

1. Frader J: Difficulties in providing intensive care. Pediatrics, 64:10, 1979.

2. Waller DA, Todres ID, Cassem NH, Anderten A: Coping with poor prognosis in the pediatric intensive care unit. Am J Dis Child, 133:1121, 1979.

3. Youngner S, Jackson DL, Allen M: Staff attitudes towards the care of the critically ill in the medical intensive care unit. Crit Care Med, 7:35, 1979.

4. Clark H, Marshall C: Definite reference and mutual knowledge. Presented at Sloan Workshop on Computational Aspects of Linguistic Structure and Discourse Setting. University of Pennsylvania. May 24-27, 1978.

5. Donnellan K: Speaker references, descriptions and anaphora, in Cole P (ed): Syntax and Semantics, Vol. 9, Pragmatics. New York, Academic Press, 1978.

6. Gazdar G: Formal pragmatics for natural language: implicature, presupposition, and logical form. Ph.D. Dissertation, University of Reading, 1976.

7. Gazdar G: Pragmatics. New York, Academic Press, 1978.

8. Hawkins J: Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction. Atlantic Highlands, New Jersey, Humanities Press, 1978.

9. Karttunen L: Presuppositions of compound sentences. Linguistic Inquiry, 4:169, 1973.

10. Karttunen L: Presupposition and linguistic context. Theoretical Linguistics, 1:181, 1974.

11. Karttunen L, Peters S: Conventional implicature, in Oh C-K, Dinneen D (eds): Syntax and Semantics, Vol. 11, Presupposition. New York, Academic Press, 1979.

12. Kempson R: Presupposition and Delimitation of Semantics. Cambridge, Cambridge University Press, 1975.

13. Prince E: On the function of existential presupposition in discourse. Proceedings of the Chicago Linguistic Society, 1978.

14. Prince E: On the inferencing of indefinite--<u>this</u> NPs. Presented at the Sloan Workshop of Indefinite Reference. University of Massachusetts, December, 1978.

15. Lakoff G: Hedges: a study in meaning criteria and the logic of fuzzy concepts. Proceedings of the Chicago Linguistic Society, 8:183, 1972.

16. Lakoff R: The pragmatics of modality. Proceedings of the Chicago Linguistic Society, 8:229, 1972.

17. Prince E: The syntax and semantics of NEG-raising, with evidence from French. Languages, 52:404, 1976.

18. Bryant GD, Norman GR: The communication of uncertainty. Proceedings, 18th Annual Conference on Research in Medical Education, Washington, D.C., November, 1979. (Published by the Association of American Medical Colleges)

295

SYMPOSIUM

EXPLICIT DEFINITIONS OF COMPETENCE FOR GRADUATE MEDICAL EDUCATION: WHAT, HOW AND SO WHAT?

Chairman: Philip G. Bashook, Ed.D.

Participants: Jerry D. Gates, Ph.D.
John S. Lloyd, Ph.D.

Discussants: Fredric D. Burg, M.D.
Leslie J. Sandlow, M.D.

EXPLICIT DEFINITIONS OF COMPETENCE FOR GRADUATE MEDICAL EDUCATION: WHAT, HOW AND SO WHAT?

Chair/Organizer: Philip G. Bashook, Ed.D.
Michael Reese Hospital and Medical Center
Educational Development Unit

Objectives

1. What written definitions of expected competencies have been developed?

2. How good are these written documents for their intended purposes?

3. How are they currently being used and what are future expectations?

4. What issues have been raised concerning using these documents for:

    - Certifying and recertifying practitioners
    - Planning residency programs
    - Resident evaluation
    - Setting resident expectations
    - Providing the public with standards for judging physician performance

5. Should these efforts be continued, enhanced or dropped?

Over the past five years many specialty examination boards in conjunction with specialty associations, and residency training programs have been developing written descriptions of expected competencies. Nearly all of these efforts have three common elements: they take an inordinate amount of professional energy and require years to complete an initial document; they represent a kind of concensus about the knowledge, skills and attitudes in the respective discipline(s); and they have generated controversy within the medical education community about the benefits, frustrations and implications of having explicit written expectations for members in a particular medical field.

The symposium is organized into two presentations and two discussions plus opportunity for extensive audience interaction. The chairman will give a five-minute introduction which will contain: an operational definition of terminology, state the symposium objectives and provide a context for considering the presentations and discussions. Both presentations will be 15 minutes long and address questions One, Two and Three. The presentation by Dr. Gates considers the use of explicit competencies within residency training programs. The presentation by Dr. Lloyd considers their use from the certification perspective.

Each discussant will have 15 minutes to address questions Four and Five. Both discussants agree that explicit descriptions of competencies are an essential ingredient in certifying competence and operating effective residency training programs. Dr. Burg contends that these efforts have been a success and should be greatly enhanced. He cites as evidence current certifying exams to evaluate practitioners considers not just knowledge as in past examinations, but other issues of competence. He also contends the process of defining competencies demonstrated a mechanism which can be useful for conceptualizing many other aspects of residency training and evaluation. Dr. Sandlow contends that nationally produced competencies are not being adopted by individual residency programs. He suggests either the competencies are not appropriate to the residencies even though the specialty boards and others use them for national examinations, or the methodologies employed in developing them do not use effective socio-political techniques to assure adoption by the residency programs.

Following the discussants remarks there will be open discussion among the symposium panelists and audience.

EXPLICIT COMPETENCIES FOR RESIDENCY TRAINING PROGRAMS

Jerry D. Gates, Ph.D.
Educational Development Unit
Michael Reese School of Health Sciences

The quality of graduate medical education has become one of the major focal points in assuring a competent medical profession. As of 1979 there were 4,630 individual residency programs representing 36 different specialties, which translates into over 64,000 residents in training.[1] The residency phase is no longer a time to just acquire a proficiency in a technical field, but rather it is a time in which the new physician develops the mental set and specialty skills for maintaining a competent practice.[2] Since it is during the residency phase that a physician shapes his/her career, formulates significant attitudes, and begins to develop patterns which will mold the practice of medicine for that person, it is universally accepted that the quality of graduate medical education should be of prime concern. Today only a few residency programs have explicitly described their training in order to assure the competence of their graduates to practice medicine.

The focus on competencies during residency training has taken several avenues, such as certification of graduates, accreditation of programs, establishing practice competencies, and so on. While no one would argue against maintaining the highest quality residency training program as possible, to establish the grounds of what constitutes quality is difficult. Quality residency training depends upon: 1) the number of faculty, 2) the clinical and teaching abilities of the faculty, 3) the number and diversity of patients, 4) the quality of the supervision, and 5) the expected roles residents are to assume. But these factors are not, alone, sufficient to assure that the graduating resident will be a competent physician. A key factor in assuring the quality of residency training, as mentioned above, is assuring that the residents are competent physicians when they finish their training. The question is, however, competent in what?

Traditionally, graduate medical education, as a whole, has left the question of competency to the specialty boards. Without going into all the arguments pro and con at this time, let it be sufficient to say that this approach short changes each unique residency program. It should be each program's responsibility to assure the quality of its own graduates. In order to do this a program needs to establish criteria for successful graduation. They can do this better than the specialty boards since they are tuned in to the medical needs of their community, while specialty boards are trying to modulate among all practice settings possible.

Before these competencies are used to measure the quality of each graduate, the competencies have to be first defined and then accepted. This presentation will address how some residency programs have come to grips with the definition issue.

Today, only a few have attempted to define competencies for their own programs. Since there are so few, it is extremely important to take a closer look at some of these defined competencies. This presentation will focus upon a selected sampling of competencies which have been developed. The presentation will be a comparative analysis of these select written competencies. Five major points will be discussed which are briefly described below.

First, the presentation will analyze the major purpose and/or intents of these selected competencies. Three major purposes will be discussed: some competencies were developed to judge the resident in a summative fashion, other competencies were intended during training, and finally, the intent of other competencies were to be used by the teaching faculty in directing them towards relevant teaching.

Second, will be a comparison of the different methodologies which were utilized in the development of written competencies. The use of experts in formulating a consensus was the method most often used. However, other methodologies were used such as critical incident and surveys of practicing professionals. How and by whom the competencies were developed will be examined in light of how this may have influenced the nature and usefulness of the written competencies.

Third, this presentation will thoroughly examine the content, nature, structure and characteristics of the written competencies. Some residency training competencies are written to present expected practice behaviors, other competencies are merely samples of the level of content to be mastered, while other competencies represent a list of specific "behavioral objectives." And still others are explicit descriptions of the "knowledge" content of the practice of medicine. Of special emphasis will be analysis of the degree of explicitness of the written competencies and their usefulness.

Fourth to be addressed in the presentation is the actual use of the defined competencies. The use is, of course, directly linked to the three factors mentioned above. The degree of acceptance and effects will be specifically identified. Factors which may have discouraged the adoption or full use of the written competencies will be addressed.

Finally, this presentation will open the question of whether or not competencies developed for training residents are reflective of the actual practice of medicine. That is, do these competencies cover the appropriate depth and breadth of the specialties they are intended to cover, and are they sufficient to assure competent graduates.

After addressing these five major issues, the presentation will conclude by answering the all important question: "Are they really worth the effort and has it made a difference in the training of residents to practice medicine?"

References

[1]Directory of Residency Training Programs, 1979-80. Chicago: American Medical Association, 1980.

[2]Cooper, J.A.D. Undergraduate Medical Education. In Advances in American Medicine: Essays at the Bicentennial, Vol. I, Parcell, E.F. and Bowers, J.Z. (eds.). New York: Josiah Macy, Jr. Foundation, 1976.

EXPLICIT DEFINITIONS OF COMPETENCE DEVELOPED BY SPECIALTY BOARDS

John S. Lloyd, Ph.D.
Director, Education and Research
American Board of Medical Specialties

Along the continuum of medical education specialty boards are situated at the nexus between graduate medical education (GME) and continuing medical educaton. Specialty boards provide an external evaluation of GME at two points: they evaluate GME on the input side (accreditation) and on the output side (certification). Specialty board certification has been awarded to those physicians who have met certain "credentialing" requirements (for example, a specified number of years of residency training) and who have achieved a passing score on the boards' certifying examinations. Thus, certifying boards assess the special competence of individual physicians at the end of GME. Recently, specialty boards have established procedures for measuring the continuing competence of physicians they have certified.

Before measuring competence it is necessary to define it with reference to the particular specialty involved. Although all specialty boards develop their examinations according to some definition of the specialty, these definitions vary in their detail and explicitness. The more explicit and detailed of these documents will be called here "definitions of competence." Those definitions of competence which have been developed by (or for) specialty boards, and which have been published, will be the subject of this presentation.

The American Board of Orthopaedic Surgery perhaps was the first specialty board to develop an explicit definition of competence in 1965. This definition was developed to improve the board's certifying examination procedures as well as the effectiveness and efficiency of residency training programs in that specialty. It was based on a survey of critical incidents from over 1,000 orthopaedic surgeons from the United States and Canada. These critical incidents were classified into nine areas of competence and 94 subareas. These areas and subareas were arranged in outline form, followed by examples of effective and ineffective (i.e., critical) incidents. The resulting document was published by the board and the American Institutes for Research (in the Behavioral Sciences) and was entitled "Critical Performance Requirements for Orthopaedic Surgery."

During the ensuing 15 years, five more definitions of competence have been developed by (or for) the following specialty boards:

- American Board of Nuclear Medicine
- American Board of Pediatrics
- American Board of Psychiatry and Neurology
- American Board of Internal Medicine
- American Board of Emergency Medicine

Each of these definitions of competence will be described along with its expected and actual use. Possible future uses of definitions will be considered.

<u>SYMPOSIUM</u>

COST AWARENESS EDUCATION AND PRACTICE OF MEDICINE:
SOME NATIONAL AND INTERNATIONAL PERSPECTIVES

Chairman:	Jack L. Mulligan, M.D.
Organizer:	Mohan L. Garg, Sc.D.
Participants:	Evert Reerink, M.D., Ph.D. George A. Schlichte, Ph.D. W.M. Kleinberg, M.D. Mohan L. Garg, Sc.D. W.A. Gliebe, M.A.
Discussants:	Joseph Gonnella, M.D. William Sodeman, Sr. Mohan L. Garg, Sc.D.

SYMPOSIUM

COST AWARENESS EDUCATION AND PRACTICE OF MEDICINE: SOME NATIONAL AND INTERNATIONAL PERSPECTIVES

Organizer: Mohan L. Garg, Sc.D.
Professor, Dept. of Medicine and
Director, Dept. of Cost Containment & Evaluation
Medical College of Ohio
C.S. # 10008
Toledo, OH 43699

The purpose of the symposium is to present approaches to cost awareness and quality assurance education for practicing physicians, especially as they apply to continuing medical education.

During the last two AAMC annual meetings, symposiums have presented curricula for cost containment in the area of undergraduate and graduate medical education. The time has, therefore, arrived to extend the education programs to continuing education. There are more than 300,000 practicing physicians in the U.S. who are directly responsible for nearly 70% of the personal health care expenditures. Some medical specialty boards require a certain amount of minimum credit hours of continuing medical education for recertification and many states have CME requirements for relicensure. However, presently we have no specific programs which address the issues of cost awareness education. This symposium will present three different approaches, all dealing with cost awareness education for practicing physicians.

Chairman: Jack L. Mulligan, M.D.
Professor
University of Missouri at Kansas City School of Medicine
2411 Holmes
Kansas City, MO 64118

ABSTRACTS

QUALITY ASSURANCE AND COST CONTAINMENT: FEEDBACK TO MEDICAL EDUCATION

E. Reerink, M.D., C.B.O.
National Organization for Quality
Assurance in Hospitals
Utrecht, The Netherlands

In 1976, the Dutch National Organization for Quality Assurance in Hospitals was set up by the professional organizations, the National Hospital Organization, financiers of health care and the Ministry of Health. Its primary goal was to assist, guide and evaluate physicians nation-wide in quality assurance activities in hospitals.

To this effect, medical care evaluation studies were initiated, implemented and assessed. These studies focused on effectiveness and efficiency of medical care (cost containment) as well. Also conflict management was included in these studies.

One hundred and forty four of these studies were analyzed for their implications for medical education, both undergraduate and graduate (specialist training). In some hospitals, multiple choice tests were administered to physicians to measure their knowledge about the management of certain medical problems. Support can be given to the findings of Williamson that the primary area for concern is not the lack of knowledge in practicing physicians, but more their lack of compliance to follow rules.

The implications for feedback to medical education at the undergraduate and graduate level on the basis of these findings and the possibilities for reconstructions of medical education in the Netherlands will be discussed. The fact that undergraduate medical education is in the hand of the universities and graduate training is in the hands of the medical profession may prove both a stumbling block to change and a unique chance to bring education and practice closer to each other.

A MODEL FOR COST/BENEFIT ANALYSIS IN CLINICAL DECISION MAKING

George A. Schlichte, Ph.D.
Health Care Planning & Development
Blue Shield of Mass, Inc.
Boston, MA 02106

"Strengthen price consciousness, or expand regulatory measures" are alternatives defined by the American Medical Association sponsored Commission on the Cost of Medical Care. Says this AMA Commission, "The most promising approach to cost containment, in the presence of insurance, is strengthening price consciousness." Administrative practices, regulations and rate setting all may be classified as bureaucratic ways to control expenditures for medical care. Essentially, these bureaucratic ways are systems of rules. Rules are essential for order, but have inherent limitations; like the weave of a fisherman's net, they have no control at the loopholes and catch only the larger fish. On the other hand, the decision making ability of the participating physician is without these loopholes; it embraces the entire practice of medicine and is directed by the needs of the individual patient. The objective of this paper is to present one way to make it possible for physician decision making to help contain expenditures for medical care. In cost accounting terms, the method would be known as "Project Costing"; each procedure becomes a "Project," thereby allowing for variations between individual cases. This method can not replace budgeting and forecasting; its purpose is to "strengthen price consciousness" of practicing physicians.

The model is a way to educate physicians in the relative costs and benefits of alternatives which deliver equivalent medical care. The cost perspective is that of the payor for medical care and the benefit considerations are those of the patient. Cost/benefit analysis take place between the patient and the physician; it is carried out through physician decisions during the treatment of illness. Hence, the task is to develop a conscious awareness by physicians of the equivalent alternatives and their relative costs.

Two kinds of physician decisions are relevant; they are the choices between alternatives in the course of treatment and the orders for the number of units of service within the chosen alternatives. Examples of alternatives are those for medical instead of surgical treatment and inpatient versus outpatient care. Examples of units within an alternative are length of stay, the number and kinds of professional support services, equipment and supplies. Initially, cost/benefit analysis will address the number and type of units within a procedure.

A first requirement is to identify the dollar costs associated with the various combinations of units within alternatives. The basic instrument for this purpose is the insurance claims submitted by both the hospital and the physician. Information contained on claims lists and types of service, their number and their respective charges. These data can be arrayed according to each attending physician's cases in a diagnostic grouping over a given period of time. Within a diagnostic grouping, data can be sorted by procedure. Out of such a display, calculation of medians and development of scatter diagrams can compare physicians on the dollar costs of various items which make up the total costs of all units within a procedure. In short, patterns of cost by procedure can be identified for individual attending physicians and each physician will have a unique location on a grid designed to allow cost comparisons between physicians. Patterns are used, rather than individual cases, because the objective is to analyze consistent cost behavior for a procedure, rather than to correct for variations from rules.

One example is cataract operations in nine non-teaching hospitals in Maasachusetts. An analysis of median charges for ancillary services showed a variation between 23 physicians of $990 for 135 patients under 65 with no complications. Ranges for physicians within the same hospital went from a low of $10 to a high of $463.

COST AWARENESS BEGINS AT HOME: SOME EXPERIMENTS IN CONTINUING MEDICAL EDUCATION

W.M. Kleinberg, M.D.
M.L. Garg, Sc.D.
W.A. Gliebe, M.A.
Medical College of Ohio
Toledo, OH

The Northwest Ohio Cost Awareness Project (NOCAP) is an educational program for practicing physicians. Developed and conducted by the Dept. of Cost Containment & Evaluation at the Medical College of Ohio, NOCAP attempts to teach cost awareness through selected methods in as many hospitals as possible in Northwest Ohio. It seeks to demonstrate that physicians are concerned about high health care costs and are trying to do something about them. The program is supported by a grant from the National Fund for Medical Education and has the cooperation of the Ohio State Medical Association.

The first objective of the program is to develop, test, and evaluate several cost awareness methodologies that can be readily integrated into currently existing continuing medical educational programs in the twenty county region of Northwest Ohio. The second objective of the project is to create cost awareness programs that, with the support of the Ohio State Medical Association (OSMA), can be implemented in CME programs throughout the state.

The hospitals in the Northwest Ohio region serve as primary contacts for planning and implementing these cost awareness educational programs. As many hospitals as are interested are invited to participate. Each program will be designed to meet the needs of the particular hospital. Where none exist and interest is shown, programs may be prepared for staff meetings of the particular hospitals. Presentations might also be made at meetings of county medical societies if that is found to be a suitable forum by local groups.

Three alternative types of involvement are currently available to the cooperating hospital. As shown in figure 1, options A and B are derived from results of a medical care evaluation (MCE) study conducted in the hospital. The MCE is generally designed to identify weak points in the patient care delivery system so that attention could be focused on their correction. Clinical topics for study, such as congestive heart failure is first agreed upon by hospital representatives and the project staff. Data elements specified by the project staff are then gathered for a selected number of patients. The data are then processed and analyzed by the project staff, with the results presented to hospital representatives. At this point options A and B emerge.

Option A presents the cost analyses through special programs to the medical staff. Other additional sessions can be presented as necessary. The programs are evaluated in two ways, by the audience at the presentation, and by a later MCE.

Option B uses the results of the MCE to present special topical programs with cost components included. Such programs will also be evaluated as above, by the audience at these presentations and by a later MCE.

Option C entails integrating cost awareness concepts into traditional educational techniques. These approaches include case presentations (or clinical-pathological-conferences) and topical presentations.

NOCAP will, hopefully, result in a significant number of practicing physicians exposed to these techniques. The project is likely to increase the pool of community physician educators who are familiar with and know how to present cost awareness concepts in ongoing educational activities. Other professional organizations will hopefully support this experiment by creating various forums for public dissemination of the experiment's development.

SYMPOSIUM

EDUCATIONAL MODELS IN PRIMARY CARE

Chairman: Edward A. Wolfson, M.D., M.P.H.

Organizer: Lawrence P. Tremonti, M.D., F.A.C.P.

Participants: Daniel L. Moser, M.S.
Craig Booher, M.D.
C. Kent Smith, M.D.
Robert H. Seller, M.D.

Discussants: W. Barry Biddle, Ph.D.
Lawrence P. Tremonti, M.D., F.A.C.P.

Symposium

EDUCATIONAL MODELS IN PRIMARY CARE

Organizer: Lawrence Tremonti, M.D., Clinical Campus, State University of New York, Upstate Medical Center, Syracuse, New York.

Chairman: Edward A. Wolfson, M.D., M.P.H.

Participants: W. Barry Biddle, Ph.D.
Craig Booher, M.D.
Daniel L. Moser, M.S.
Robert H. Seller, M.D.
C. Kent Smith, M.D.
Lawrence P. Tremonti, M.D.

Objective:

It is the purpose of this symposium to provide representatives from four different institutions the opportunity to contrast different undergraduate educational models for teaching primary care and to address issues generic to these models.

The objectives include:

1. To acquaint the audience with current models of primary care instruction.
2. To cite successful strategies inherent to each model.
3. To identify resources required for implementation of the model.
4. To contrast outcomes in terms of student career choice.
5. To review major problmes inherent in each model.

310

Educational Models in Primary Care

In "A Manpower Policy for Primary Health Care" study, the Institute of Medicine has defined the essentials of Primary Care as accessibility, comprehensiveness, coordination, continuity, and accountability. Most of the educational offerings in long established medical schools have attempted to teach individual aspects of primary care using the educational format of traditional discipline clerkships. Because of existent faculties and clinical settings, the traditional programs have generally been incomplete educational offerings relative to primary care essentials.

During the 70's, considerable impetus toward the development of educational innovations in primary care instruction has been provided by the identification of specialty and geographic maldistribution, the advent of family medicine, and the development of more than 40 predominantly public community-based medical schools. These efforts have been encouraged by our national priorities and societal needs relative to primary care. Medical schools have adopted a variety of approaches in order to influence student's career choice such that at least 50 percent of graduates would enter a primary care field (family practice, internal medicine, and pediatrics). The majority of approaches which have been implemented fall into four major formats. They include:

1. The Intense Clerkship Model in primary care embodying the role models and clinical experiences of a primary care physician. This model consists of 4 to 12 weeks of intense concentrated effort within Family Practice with a focus on the family and the community.

2. The Longitudinal Model of ambulatory primary care experience within a family physician preceptor's office or a primary care ambulatory setting. This model assigns students to specific half day assignments every week for a prolonged period up to three years within a primary care office setting.

3. The Tracking Model whereby students elect a career pathway early in their medical school experience. This model requires an early decision frame on the part of the student rather than exposing all students to the same program.

4. The Integrated Primary Care Model which attempts to intersperse various aspects of primary care throughout the medical student's four year curriculum. This model allows the students opportunities to explore issues in primary care, to render some experience in a limited fashion, and to provide exposure to family physician role models in various aspects of the student's education.

The symposium will provide not only information relative to the current models of primary care, but also an opportunity to discuss the merits of each of the models in light of the resources and the outcomes as determined by the student's choice of graduate training programs. The objectives of the symposium are:

1. To acquaint the audience with current models of primary care instruction.

2. To cite successful strategies inherent to each model.

3. To identify resources required for implementation of the model.

4. To contrast outcomes in terms of student career choice.

5. To review major problems inherent in each model.

The format of the symposium will consist of short 10 minute presentations from each of the major participants with respect to the model conducted within their own institution, the required resources, the existent problems, and the outcomes of the students having completed the program. Following the presentations, there will be two respondents on the panel who will address the appropriateness of the models relative to the essentials of primary care and the educational design and format of the models. The chairperson will solicit additional questions to engage the audience in discussion of the material presented.

It is anticipated that the presentations by the symposium participants will provide a variety of educational options which can be brought back to the audience's institution as well as provide resource persons to assist them in defining or refining the development of primary care programs within the institutions represented in the audience. The presentations will cite successful institutional strategies, address the issue of required primary care instruction, and explore how and by whom these programs should be conducted.

A REQUIRED CLERKSHIP IN FAMILY MEDICINE AND COMMUNITY MEDICINE

Daniel L. Moser, M.S., Education Specialist, Center for Educational Development, Eastern Virginia Medical School, Norfolk, Virginia 23501.

Christopher M. Buttery, M.D., Director of Public Health & Welfare, Corpus Christi, Texas 78408.

Eastern Virginia Medical School was founded in 1973 with a strong community and primary care orientation. It is funded from local resources to a large extent and uses over twenty-five local hospitals for teaching instead of a university hospital. As part of the emphasis on primary care, an eight week Family Medicine clerkship has been required for all students.

Design and Functioning of the Clerkship

The Family Medicine clerkship has two focuses, Family Medicine and Community Medicine. For both pedagogic reasons and the local situation, the clerkship was designed with these two focuses in a complimentary fashion. The clerkship is made up of five half days in a Family Practitioner's office, a half day each week in an alcohol abuse program, in an emergency room, and in a public health program, and two half days in seminars covering family and community medicine issues.

The sessions in physicians' offices are concerned primarily with the care of ambulatory patients. Students are assigned to gather medical histories and complete physicals, make diagnoses, and prepare plans of care on several patients daily. Additionally, they accompany the physician while he sees patients and participates in the patient's care, sometimes with the physician observing their skills and sometimes observing the physician's care of the patient. These sessions cannot be standardized experiences, but care has been taken through direct faculty contact and data gathering on patients seen by students to assure minimum standards of responsibility and comprehensiveness of experiences. The office-based curriculum's main objectives will be distributed.

The public health department practicums provide information and experience in the entire range of community support available to the family practitioners and their patients. They include all aspects of public health including administration, vital data collection, public health law, categorical clinical programs, home health visiting, health education, and environmental health programs.

As a special public health emphasis, students also participate in an alcohol and substance abuse program. This involves a student in the care of chronic alcoholism including observation of role models and the actual delivery of care, under supervision, to these patients. The emergency room experience gives them direct experience in the care of emergencies often handled by Family Practitioners in their offices and at the hospital.

From national test results, we have found that our students score above the national average consistently on the public health section. In addition, we have seen an increase in the number of graduates who are doing residencies in Family Practice from 13 percent of the class for the first class to 37 percent for the fourth class which graduated in 1979 and 17 percent for the latest class. (The latter decreased rate due in part to difficulties experienced with larger classes.)

In the follow-ups with graduates of the school, residents have overwhelmingly felt the actual experience in physicians' offices was the most valuable aspect of the clerkship. Interestingly not always for its positive results, a few students have noted it made them sure they didn't want to go into Family Practice. Assuming the experience was appropriate, this effect although not desired is probably for the best of the student.

From our experience the size of the group of students has been crucial to the success of the clerkship. As we have approached and passed our maximum desirable number for the resources available, the experience has started to deteriorate for the students. Unlike programs that use the residency site exclusively for the Family Medicine clerkship, our system has still not found an adequate role for the residency sites or residents in our clerkship. In regards to motivating non-Family Medicine oriented students, our record speaks for itself; they have been given a varied, intensive, scholarly experience in Family and Community Medicine and a large majority have appreciated it.

Conclusions

A combined family and community medicine clerkship is feasible. It appears to have had a significant effect on choice of residency by students at this school, as well as on the development of clinically oriented preventive medicine skills that relate to daily medical practice. In addition, the clerkship is highly rated by students. This form of primary care clerkship is possible with minimal financial support and with the commitment of volunteer faculty in the community.

THE ROCKFORD FILES: A LONGITUDINAL PRIMARY CARE MODEL UPDATE

Craig Booher, M.D., Associate Dean, Rockford School of Medicine, University of Illinois College of Medicine.

The Rockford School of Medicine provides three years of clinical medical education to students who have completed one year of basic medical sciences in the Schools of Basic Medical Sciences at Chicago or Urbana. Unique within the Rockford School of Medicine's curriculum was a structured Primary Care Experience with a two-fold intent.

The Primary Care Experience which represents approximately 20 percent of the curriculum was designed as the ambulatory care phase of the Rockford School of Medicine's curriculum. The Primary Care Experience (beginning after completion of the Preclinical Experience in January of the sophomore year) requires that the students to be assigned to one of four community health centers for two-half days per week for the remaining 2½ years of their undergraduate medical education. Through the development and maintenance of their patient panels, students are expected to acquire under close supervision the skills and responsibilities associated with primary care medicine.

Four primary care centers were designed and developed in small communities which previously were determined to have insufficient physician coverage. Located within 34 miles of Rockford, the centers are structured as a fee-for-service group practice and serve as the primary care educational vehicle for the students. As such, they are the very cornerstones of the ambulatory curriculum emphasizing continuity of care. A brief history of the development of each center will be discussed, covering in general, architectural plans, and building costs. The resource requirements for teaching primary care to undergraduate medical students using the Rockford model will be discussed with particular emphasis on quanitative and qualitative aspects of the patient panel, the required faculty, staff, and the optimum number of students.

In order to answer the question relative to the level of success with the Rockford Model, a study was undertaken after six years of operation intended to examine the effectiveness of the Rockford School of Medicine Undergraduate Primary Care Experience. Three hypotheses were tested:

1. The education of students in the community health center will, in conjunction with other elements of the educational program, increase the likelihood that the student will go into primary care residency/practice.

   The vast majority (88.3%) of past and present students entered the Rockford School of Medicine with the intention of pursuing careers in primary care medicine. The data indicated that a slight shift away from primary care intentions began almost immediately during the sophomore year and increased through the junior and senior years. Minimally, nearly 20 percent of the students abandon their primary care career intentions while at R.S.M. and 61 percent of graduates enrolled in primary care residencies.

2. The education of students in the community health center will, in conjunction with other elements of the educational program, increase the number of students who desire to practice in rural or small town settings.

   a. It appeared that the desire to practice in rural locations diminished throughout the undergraduate years and is related to specialty choice decision.

   b. The vast majority (83-96%) of the students expressing intentions to establish practices in rural areas also planned careers in primary care medicine, while those planning practices in urban areas were almost equally divided in their primary versus non-primary specialty choices.

   c. It was seen that students from rural areas, upon entrance to R.S.M., hold rural/primary care intentions nearly twice as frequently as students from urban ~~areas~~; however, both groups decline such intentions over the ensuing years.

3. Medical students will rate their community health centers instruction, instructors, and health center learning environment as more influential in their careers than other elements of the curriculum.

   Personal preferences, namely geographic locations and input from the spouses, were the most important considerations in the particular residency selection of R.S.M. graduates. However, when specialty choice was examined, influences directly related to the Rockford School of Medicine and/or professional experiences resulting there from were found to be more influential.

Data gathered from each of the graduating classes since 1975 reporting career choice, graduate training, and subsequent practice sites will be discussed. It is of interest that we are experiencing our first influx of former graduates who have completed their residency training and who have chosen to come back to Rockford to practice because of the opportunity to become a physician teacher.

A FAMILY PHYSICIAN PATHWAY MODEL AND ITS EFFECT ON CAREER CHOICE

C. Kent Smith, M.D., Department of Family Medicine, University of Washington School of Medicine.

The beginnings of a systematic effort to train primary care physicians at the University of Washington School of Medicine can be traced to a major curricular revision which oocurred in 1968. Major goals for this restructuring were to increase the number of graduates specializing in primary care and the number wishing to practice in underserved areas. The vehicle selected for achieving these goals was the development of a department of Family Medicine.

The educational model which was evolved over the past twelve years can be broken down into six elements. Our progress in pursuit of these goals can be largely attributed to these six components which are described in the following paragraphs.

1. Full departmental status, including number of faculty, faculty rank, committee representation, and curricular space. Conferral of departmental status, achieved in 1971, was considered essential, both to establish visibility and credibility with students and to provide the capability to develop comprehensive educational programs at the undergraduate and graduate levels.

2. A strong advising system providing family physician role models early in the medical student's career. Such a system was initiated with the 1968 curricular change which established the Family Physician Pathway as one of four available options and required students to select a pathway and advisor midway through their second year. As the number of students selecting this pathway and the number of full-time faculty have increased in concert over the years, the advising system has become increasingly important in building and sustaining student interest in Family Medicine.

3. Course representation in each year of the curriculum. Students interested in primary care must be provided curricular opportunities to learn the principles of family medicine and receive periodic reinforcement of this training throughout their undergraduate careers. This department has for a number of years offered a first-year introductory preceptorship and a second-year nine-month continuity preceptorship, both taught in the offices of Seattle family physicians. The major course in the department, a community clerkship, annually provides training for 100 third- and fourth-year students at eight sites located throughout the four-state area served by the University. In addition, an advanced preceptorship is offered in the fourth year. Full curricular coverage was achieved in 1978 with the introduction of a third-year basic clerkship taught at nine residency training sites affiliated with the department.

4. Representation in the required curriculum. It is felt that exposure to family medicine is beneficial to all students regardless of their specialty interests. Faculty members from this department chair a first-year course in the Introduction to Clinical Medicine taken by all students. Although not formally a part of the Family Medicine curriculum, this course provides a strong emphasis on the principles of Family Medicine. In addition, we are in the process of increasing the capacity of the community clerkship described above to accommodate the entire third-year class.

5. Training in medically underserved areas. If we wish to encourage the interest of students to practice where they are most needed, it is important to provide them with learning experiences at these locations. This department offers the six-week community clerkship taught at eight primarily rural sites. The advanced preceptorship is taught at five sites established by NHSC and IHS in underserved areas. In addition, several of the residency training sites used by the third-year clerkship are located in Seattle's inner city area.

6. Affiliation with residency training programs. Such programs provide valuable undergraduate training sites, exposure to resident role models, and attractive, visible options for students when they make their career choices upon graduation. This department is fortunate in having

established a strong affiliation network with eight effective programs which provide training for 141 residents each year.

In terms of outcome measures, our program must be considered quite successful. Almost 50 percent of each second-year class since 1970 has selected the Family Physician Pathway in preference to the other three options. Of this number, 73 percent or almost three out of four have entered graduate training for family practice. For the school as a whole, approximately one third of all graduates now enter training for family practice. This compares to an estimated five to fifteen percent entering general practice in the early 1960's.

The model described above provides an overview of the steps taken by one department to develop a comprehensive program of education in primary care. It fails to give recognition to the hard work, talent, and dedication of the many university and clinical faculty who participated in this development. Results to date indicate that although much remains to be accomplished, considerable progress has been made toward meeting the needs for primary care physicians in our region.

AN INTEGRATED CURRICULAR APPROACH TO THE TEACHING OF PRIMARY CARE

Robert H. Seller, M.D., Professor & Chairman, Department of Family Medicine, Medical School of the State University of New York at Buffalo.

SUNY-Buffalo is a long established school with a fairly classical curriculum as contrasted to new schools whose curriculum was created to emphasize training students for primary care. At SUNY-Buffalo, various aspects of primary care are taught by several departments in all four years of the curriculum. In the first year, students are introduced to clinical medicine and some of its primary care aspects through a clinical advisorship program. Groups of 3 or 4 students are assigned to physicians who meet with them informally several times a month. The largest number of students are assigned to clinicians in the primary care specialties. As part of this program, the students visit their physician's office and occasionally observe the physician's hospital practice. This program is designed to introduce freshman students to various aspects of clinical medicine and also to provide them with a personal, informal way of learning some health care and life-style issues that await them as practicing physicians. Students continue to meet their clinical advisors during their second year but less frequently.

The major exposure to primary care in the second year is a required course in Family Medicine. Individual students are assigned to a practicing family physician and also to a small (10 to 12 students) seminar group which meets with two full-time members of the Department of Family Medicine -- a family physician and a behavioral scientist. For eighteen weeks the students meet one half day per week alternately with their family physician preceptor and the small-group seminars. While in the family physician's office, students participate, as appropriate, in the care of the patients. This provides the students a first-hand opportunity to learn the diagnosis and management of some common ambulatory problems, as well as how family physicians deal with problems in family dynamics. The seminars are designed to focus on issues of particular relevance to family medicine and primary care, such as comprehensive and continuing care, patient complicance, chronic illness, death and dying, the use of community resources as adjuncts to clinical intervention, and family dynamics.

Although the junior year deals primarily with the customary inpatient clerkships, a small portion of each clerkship is spent in the outpatient departments. A one-month Family Medicine Preceptorship is being introduced into the junior year. Summer electives for students who have completed the junior year include a large MECO Program (which emphasized small town and rural settings) and other primary care experiences.

Although our senior year is largely elective, there are two required months, one of which is a one-month subinternship. The second requirement which was recently introduced, is one month of primary ambulatory care. The student may choose to take this month of primary ambulatory care under the sponsorship of the Department of Medicine, Pediatrics, Ob/Gyn, or Family Practice. Most of the departments satisfy this requirement by offering a rotation in the Outpatient Department. With the Department of Family Medicine, the student is usually assigned a full-time preceptorship with a family physician. Students are encouraged to take this rotation with physicians practicing in rural or semi-rural areas. The Department has worked with physicians and communities to provide housing for the students and, in some instances, funds to defray travel expenses. Students may also fulfill this requirement by working in the Family Practice Center of the Family Practice Residency Program. Approximately 50 percent of the students choose to fulfill this requirement under the sponsorship of the Department of Family Medicine even though they may not be planning a career in family practice. In addition to the required elective in primary ambulatory care, many departments offer other senior electives in the primary care aspects of their specialty.

The implementation and integration of primary care education into the SUNY-Buffalo curriculum has been a gradual one and overall has met with a high degree of student and faculty acceptance. Over the last five years there has been a gradual but marked increased in the percentage of graduates entering primary care specialties including family practice. It is difficult to tell whether the aforementioned changes are related to student selection, curricular changes, changing student interests, or societal pressures.

SYMPOSIUM

MEDICAL STUDENT EDUCATION FOR RURAL PRACTICE: INFLUENCE OF CURRICULUM AND LEARNING SITE

Chairman: Arthur Kaufman, M.D.

Participants: Arthur Kaufman, M.D.
Paul T. Werner, M.D.
Tom Cullen, Ph.D.

Discussants: Ronald Richards, Ph.D.

Medical Student Education for Rural Practice:
Influence of Curriculum and Learning Site

Arthur Kaufman, M.D. (Organizer)
Co-Director, Primary Care Curriculum
University of New Mexico School of Medicine

Objective of Symposium

A growing national concern about the maldistribution of physicians by geography and specialty has led to exploration of causes and development of remedies primarily at the post-graduate and practice level. However a contributing cause must be the urban location and subspecialty orientation of traditional medical schools, which provide students with a poor model for primary care practice, especially in rural areas.

A number of medical schools are attempting to redress this imbalance through innovations in curriculum design and training site. The experience of three such schools, each offering different approaches to a common goal, are presented.

The symposium objectives are threefold: 1) to critically examine three unique medical education programs which seek to alter either the predoctoral curriculum or the site of student learning or both, to the end that more graduates will select specialties appropriate to rural practice and rural practice sites; 2) to discover common program successes and problems and discuss their importance for planning future curricular innovations for rural practice and 3) to assess the impact to date of these programs on both the rural communities designated as training sites and on the conventional curriculums or urban campuses of their respective schools.

The New Mexico Primary Care Curriculum

Arthur Kaufman, M.D.
University of New Mexico School of Medicine

New Mexico is a large, sparsely populated state in which physicians tend to cluster in the two metropolitan areas of Albuquerque and Santa Fe. It ranks 49th in primary care physicians/1,000 population. As the sole medical school in the state, the University of New Mexico feels a special obligation to serve the state's unmet health needs.

After reviewing the meager return of medical school graduates to New Mexico, it was decided to develop an experimental curricular track, the Primary Care Curriculum (PCC), that would attempt to better train medical students for the rigors and complexity of rural primary care and employ educational strategies to encourage interest in eventual primary care practice in rural New Mexico.

PCC was established at the medical school in Albuquerque alongside the conventional track. It accepts an increasing proportion of the fixed first year class size of 73; 10 the first year, 15 the next, and 20 the third year. PCC admitted its first class in 1979.

The admissions process is designed to facilitate evaluation of the curriculum. Applicants desiring PCC can only be considered after successful admission to the medical school. The entering class is then divided into four study groups. The experimental group consists of students preferring and acceptable to PCC and randomized into it; control group 1 consists of students preferring and acceptable to PCC but randomized into the conventional track; control group 2 consists of students preferring but not acceptable to PCC and randomized into the conventional track; and control group 3 consists of students preferring and accepted into the conventional track. This study design permits program evaluators to follow students prospectively and later determine which outcomes were related to the admissions process, which to self-selection and which to the curriculum.

PCC is completely separated from the conventional track during the first two years. Students are engaged in clinical, problem-based learning in small group tutorials. Problems selected in the early months are those most commonly encountered in rural New Mexico.

After acquiring a foundation of basic and clinical science knowledge during the first eight months of medical school, PCC students have attained a formal clinical skill level somewhat comparable to that of a beginning physician's assistant. They then embark upon a four to six month rural clerkship with a primary care physician or health team during which they care for patients under close supervision and identify and study community health problems.

After this substantial dose of rural, primary care role modeling, students return to the medical school for an advanced, clinical problem-based curriculum after which they sit for the National Board of Medical Examiners Part I.

The third year consists of in-hospital clerkships taken along with conventional track students and serves as an important phase of comparison between students in the two tracks. The final year for PCC differs from the conventional track in that PCC students return to a rural New Mexico community for a three to six month rural sub-internship designed to reinforce rural, primary care practice after acquiring more advanced clinical skills and insights.

After one year in operation, PCC has focused its data collection in three important areas: 1) impact of PCC on the medical school; 2) academic and attitudinal comparison between PCC and conventional track students; and 3) impact of PCC on the rural practice communities.

PCC is highly visible and cannot isolate itself from the close scrutiny and criticism which has stimulated important improvements in the program. At times, however, criticism emanates from those whose minds are closed to the value of problem-based learning or to education away from the high technology centers. Enormous energy, so needed in developing and modifying the fledgling, experimental program, is often siphoned off into selling and defending the program within the medical school.

Yet all parties agree that the institution of an experimental curriculum has energized the faculty and stimulated interest in problems and options in medical education. Increasing numbers of faculty are requesting to participate in PCC and, most important, a growing number of basic science departments are employing problem-based learning for their conventional track students.

To compare science content acquisition between students in both tracks at the end of the first academic year, PCC employed the Quarterly Profile Examination developed at the University of Missouri-Kansas City; an eight hour, 400 question basic and clinical science comprehensive examination. As expected, PCC students scored significantly higher than conventional track students in clinical subjects. But, surprising to some, PCC students scored the same as conventional track students on the basic science portion of the examination.

To assess change in student attitudes from the beginning to the end of the first medical school year, students in both tracks were administered scales assessing attitudes toward learning. Whereas conventional track students exhibited an increased apathy toward learning and a feeling of decreased relevance to medicine of basic science material, PCC students showed no such waning interest and the perception of basic science relevance remained high.

Finally, as the first group of students start their rural clerkships, PCC has already begun to build important bridges between the medical center and the state's rural practice community. Sophisticated links to rapid information sources at the medical center library are now being assembled for PCC's rural preceptors, and the medical school is now receiving requests from rural physicians for such appropriate services as an updated list of textbooks for a community hospital library and an organized telephone consultation service linking rural physicians to University Hospital specialists.

The Upper Peninsula Medical Education Program

Paul T. Werner, M.D.
Michigan State University

The Upper Peninsula Medical Education Program is a full four-year alternative campus and program of the College of Human Medicine of Michigan State University. It grew out of the mounting need for primary care, rural health services in Michigan and located in the Upper Peninsula where such needs are grossly apparent and where local community and legislative backing for such a program was strong. A weighted admissions procedure gives preference to Upper Peninsula residents and applicants from a rural background.

To foster appropriate physician role modeling and generate attachment to rural communities, all but ten weeks of the program occur at sites in Michigan's Upper Peninsula, approximately 400 road miles north of the mother campus at East Lansing.

Curricular innovations in basic, clinical and behavioral science have been made which integrate these disciplines into patient care problems. Utilizing a patient problem-oriented learning approach, students can relate the relevant science disciplines to real life medical practice. Thus, Ambulatory Patient Care Clinics are productively woven into the curriculum from the first week of school. Consistent with the curriculum's emphasis on the importance and complexity of primary care, the last years are centered in ambulatory rather than hospital settings. Patient problems in the various clinical disciplines are viewed as they appear in the family physician's office or on the community hospital wards.

In light of the school's non-traditional approach to education and its physical remoteness from the main campus, much time and effort has been devoted to convincing main campus faculty and students alike that a non-clerkship, ambulatory, outpatient curriculum can cover the objectives of an entire medical school training program. In so doing, evaluation consisted of multiple pieces taken from another setting in a traditional curriculum applied to assess Upper Peninsula's integrated curriculum. A more comprehensive evaluation system, functioning in a continuing rather than episodic fashion and keyed to the program's expressed goals thus had to be developed.

To date, a series of examinations administered both to Upper Peninsula Medical Education Program students and main campus students during the four school years have shown comparable skill level. FLEX examinations administered to graduates of both tracks reveal equivalent performance levels on both clinical and basic science portions of the examination.

However, a more important preliminary outcome is the residency selection of the two classes which have graduated to date. Of twenty graduates, seventeen have selected residencies in primary care. Data on planned practice location is only available for the first graduating class but shows nine of ten planning a rural practice, four back in the Upper Peninsula.

The Escanaba program has affected the main campus. Being remote from traditional basic and clinical science departments, it has been able to preserve a 'purer' focal clinical problem approach to medical education than has the problem-solving track at East Lansing which has gradually drifted back to a more traditional approach. The success of the Escanaba experiment has thus led to redoubled efforts on main campus to retain the principles of problem-based learning. On the other hand, Upper Peninsula's remoteness from a large university leaves the program with little depth in basic science resources. This has remained a major problem.

Finally, the Upper Peninsula Medical Education Program has had a substantial impact upon the surrounding medical communities. The recent influx of young physicians into Escanaba, swelling the medical community from 22 to 36 since the inception of the medical school, has, in part, been attributed to the opportunity afforded community physicians to teach medical students. The impact of the medical school can also be seen in the quality of medical practice in the community. The hospital's medical staff now enjoys weekly grand rounds and continuing medical education, and community services have developed such as a hospice movement and a jail health project.

The WAMI Program

Thomas J. Cullen, Ph.D.
University of Washington School of Medicine

The states of Washington, Alaska, Montana and Idaho, the WAMI states, represent 22 percent of the land mass in the United States and contain approximately 6 million people. Two-thirds of the physicians and other health professionals in this region live in three areas including the Puget Sound region of Washington, the Anchorage area of Alaska and the Treasure Valley of Idaho. The remaining third practice in the towns and rural communities of a region stretching across five time zones.

The University of Washington in Seattle has the only medical school in the WAMI region. Because much of the land is mountainous and inaccessible, the population is sparse and widely separated.

The WAMI program, which was proposed in 1970, was designed to meet four major health challenges in the Pacific Northwest and Alaska.

1. Admission to Medical School

Beginning in the late 1960's, a progressively greater number of students applied for admission to medical school without a corresponding increase in the class sizes of medical schools. This increase was especially important for states without medical schools because state-supported schools have traditionally afforded their own residents preferential access to the positions available. Admissions of out-of-state students have been uncertain and, in the face of increased resident applicants, were expected to be curtailed.

2. Lack of Primary Care Physicians

In 1970, or the year the WAMI program was proposed, the WAMI states had too few family physicians, primary care internists and primary care pediatricians. This deficiency was especially severe in the more rural areas of the WAMI states.

3. Maldistribution of Physicians

In the early 1970's, the majority of the physicians were located in the cities and larger towns, leaving large, sparsely populated areas with too few health professionals. When the geographic maldistribution was coupled with an insufficient number of primary care physicians, a severe shortage in the remote areas of the WAMI states was evident.

4. Access to Education and Health Care Resources

Physicians in isolated communities have restricted access to educational resources. Because of this, physicians have felt isolated, and physician turnover has been significant.

In 1970, only a limited number of tertiary health care centers existed in the WAMI states. Since 18 major health care center--including a neonatology center, a burn and trauma center, an epilepsy center, and a child development and mental retardation center--existed at the University of Washington Health Sciences Center, and since there are no additional centers of these kind in the region, it seemed desirable to find a way to bring these resources to communities with health care needs.

Finally, the WAMI states did not have the fiscal resources to build and maintain additional medical schools. It was recognized, therefore, that any program designed to meet the health care and health education needs of the region could not include capital expenditures for new facilities.

There are two separate phases in the WAMI program. In the first or university phase, four universities that do not have medical schools provide the first year of the medical school curriculum to medical students. This has allowed expansion of the entering class from 102 to 175, with 10, 20 and 20 first-year positions being reserved for Alaska, Montana and Idaho, respectively.

Courses for the university phase are planned conjointly by faculty from the five universities so that a "single, region-wide" course is taught at five locations by a "region-wide" faculty. To monitor the academic quality of the program and to determine whether students are learning in the required way, a continuous and extensive evaluation program is conducted.

At the end of the first year, all students go to the University of Washington campus in Seattle for the second year of the curriculum and for the initial clerkships in the third year of the undergraduate medical educational process.

Students then have the opportunity of participating in the second or community phase of the WAMI program. In this phase, students spend six weeks in community clinical units (CCU's) working with private physicians in rural communities in the four states. House staff officers in residency programs are also assigned to the CCU's for two to six months. These experiences in family medicine, internal medicine, pediatrics, psychiatry, and obstetrics and gynecology are designed to provide a participant with an understanding of how the community is structured and functions, what roles a physician must play in such a community, and what skills and knowledge are required to deliver high-quality health services. It was hoped that with this kind of exposure, larger numbers of physicians would be attracted to careers in primary care in under-served areas.

Since the advent of the program, 403 students have been trained in the university phase of the program including 109 at Washington State University, 108 at the University of Idaho, 100 at Montana State University and 86 at the University of Alaska. Using 14 measures, including internal and external examinations, no significant differences in the performance of students have been found between the group beginning at each university phase site and those who began at home base in Seattle.

The WAMI program has trained 953 students and 198 residents in the community phase of the program, including 458 and 46, respectively, in family medicine; 72 and 79, respectively, in internal medicine; 169 and 61, respectively, in pediatrics; 57 and 10, respectively, in psychiatry; and 192 students and 2 residents in obstetrics and gynecology. Using multiple types of evaluations including Part II of the National Boards and a comprehensive examination at the end of Year III, no significant difference in performance was found in the subdisciplinary content areas between those students who had their basic clerkship experiences in the community phase of WAMI and those who had their experience in Seattle.

SYMPOSIUM

ALTERNATIVE APPROACHES TO RESEARCH ON CLINICAL REASONING

Chairman: Arthur S. Elstein, Ph.D.

Organizer: Michael M. Ravitch, Ph.D.

Participants: Michael M. Ravitch, Ph.D.
David B. Swanson, Ph.D.
Georges S. Bordage, M.D., M.S.

Discussants: Barbara McNeil, M.D., Ph.D.

ALTERNATIVE APPROACHES TO RESEARCH ON CLINICAL REASONING

Michael M. Ravitch, Ph.D.
Office of Medical Education Research and Development
Michigan State University

OBJECTIVES

In the last decade, research on physicians' clinical reasoning has had a significant influence on medical school curricula. Patient management problems and clinical problem-solving exercises are widely used as instructional devices, as well as for evaluating the cognitive skills of medical students and physicians. Simulations, derived from research on the evaluation of physicians' performance, are now being incorporated in specialty board examinations. Decision-theoretic methods have become an accepted method for evaluating patient care and for conducting policy studies of medical technology utilization.

This symposium brings together three perspectives in research on clinical reasoning. Ravitch discusses the use of chart review for studying physicians' referral policies; this research is part of a series of studies aimed at describing clinical decision-making. Bordage reports on an investigation of memory organization and representation of knowledge in medical students and physicians; his research underscores the importance of students' structure of knowledge for learning and problem solving. Swanson synthesizes conclusions from two studies: the first, a computer simulation of medical reasoning in the diagnosis of congenital heart disease; the second, an investigation of factors in the development of diagnostic expertise in pediatric cardiology. Presenters and the discussant intend to review the effectiveness of the research methods used, as well as the theoretical and practical significance of the reported findings.

A CHART-AUDIT/POLICY-CAPTURING STUDY OF CLINICAL DECISION MAKING*

Michael M. Ravitch, Ph.D., David R. Rovner, M.D.,
Arthur S. Elstein, Ph.D., Penny Jennett
Office of Medical Education Research and Development
Michigan State University

In the last decade the study of clinical reasoning has gained increasing attention as an area of research. These basic and applied studies of judgment, problem-solving, and decision making have provided information for improving medical education and medical care.

Our research group is using several methods to study questions concerning the decision making of primary care physicians dealing with obese patients. Any single method, for all its advantages, has limitations which affect the conclusions one can draw. The methods we employ include: (1) chart review (audit of medical records; (2) review and discussion with a single physician of his records of patients with possible endocrinological problems; and (3) hypothetical case vignettes used as stimuli to study experimentally physicians' decision-making regarding referral. This presentation focuses on the determinants of our decision to use chart review and policy capturing and on some of the problems encountered.

In the early stages of planning, direct observation of care was considered. This method was rejected because it was too expensive, time-consuming, and reactive: (1) Observers would have to spend many hours in physicians' offices awaiting the visits of obese patients; (2) physicians who would permit observation might be unrepresentative of those who would not; and (3) the presence of an observer might affect the care provided and the decisions made to unknown degrees. These concerns led us to choose retrospective chart review. Chart review permits unobtrusive review of a large number of decisions made in actual clinical circumstances on real patients. Unlike simulated patients and patient management problems, the method of investigation does not affect these decisions.

We analyzed the charts of 124 obese patients who had been referred to the Endocrine Clinic at Michigan State University and of 250 obese patients randomly drawn from the general Clinical Center population. Data were abstracted from charts according to a specially developed coding frame. Two criteria were used for including patients in the random sample: body weight of 15% above normal, and age greater than 2 years. Some of these obese patients had been referred for endocrine workups, but the rate of referral was one of the objects of inquiry.

The retrospective review was used to develop a statistical model of the cues and their weights that predict the referral decisions of the primary-care physicians whose patients are in the sample. This statistical model is said to capture the policy of the clinicians, hence this approach is called "policy capturing" by psychologists interested in clinical judgment and decision making. We are investigating how well such a statistical model can capture the policies of a large group of primary-care physicians.

*Supported in part by a grant from the National Library of Medicine - #LM-03396.

Chart review has certain advantages as a research technique. (1) It is unobtrusive - we need not be concerned about the effect of "observation" on physicians and patients. (2) Using existing records, it is possible to obtain data on a large number of subjects in a relatively short time, compared with the time it would take to conduct an observational study as patients with a given problem happen to come to physicians' offices. (3) There is no need to speculate on the generalizability of research, as one would with simulated patients or PMPs. The data are "real data" which have been actually collected by physicians. Our question is, "How are these data used?"

In addition to these advantages, chart review has some limitations: (1) It is unlikely that each physician in the study will be represented by an equal number of patients, so that the range of physicians' decisions will be unevenly sampled. (2) Since each patient is seen by a different primary physician and each physician has a different panel of patients, differences in referral policies cannot be systematically attributed to patients or physicians. (3) Physicians may fail to record relevant information on the chart, although it is used in decision making; for example, it seems likely that the socioeconomic status of the patient is an important determinant of medical decisions (Barr, in press; Eisenberg, 1979), but social and psychological information seldom appears in the patient record. (4) Chart review has two other sources of measurement error: (a) physicians may make mistakes either in observing or recording findings, and (b) chart auditors may encode information from records inaccurately.

COGNITIVE REPRESENTATION OF MEDICAL KNOWLEDGE:

Categories and Prototypes

GEORGES BORDAGE, MD, MS
Bureau de pedagogie medicale
Faculte de medecine
Universite Laval

How is medical knowledge cognitively represented? Are different representations related to different learning outcomes or clinical performances? No decisive answers exist to these significant questions about the structure of memory and its role in clinical reasoning.

The Problem: The incentive to study memory in its medical and clinical context comes from common conclusions of investigators of the reasoning process in general and of clinical reasoning in particular. Elstein, Shulman and Sprafka (1978), as well as Barrows and co-workers (1978) and Kassirer and Gorry (1978) have shown that in dealing with selected medical diagnostic tasks, physicians differ very little in the strategy they use to solve problems. This clinical method is described by Elstein as "a hypothetico-deductive process summarized in a four-stage general model of medical inquiry that calls attention to cue acquisition, hypothesis generation, cue interpretation, and hypothesis evaluation" (1978, p. 277). Not unlike DeGroot's (1965) conclusion in his studies of chess players, Elstein concludes: "the differences between experts and weaker problem solvers are more to be found in the repertory of their experiences, organized in long-term memory, than in differences in the planning and problem-solving heuristics employed" (1978, p. 276).

Concurrent to these studies of medical reasoning, a more diverse set of investigators directly addressed the topic of memory organization and representation in medicine (Kleinmuntz, 1967 and 1968; Wortman, 1966, 1971, 1972 and 1975; Schwartz and Simon, 1971 and 1972; Allal, 1973; Pauker et al., 1976; Zacks and Sprafka, 1976). As is reflected in the various types of representations evoked (lists, hierarchies, networks), their conclusions cannot be so clearly grouped into a common consensus.

The Rationale: Given the unconclusive state of the art in the study of memory in medicine, broader answers are sought in more general studies of human reasoning and memory as found, for example, in cognitive psychology and psycholinguistics. Research in these areas, although active, is still in its infancy. Although differing in their conclusions, most investigators agree on the importance and critical role of categorization in the cognitive representation of knowledge.

Based on Gestalt theories of the first half of the century (Wertheimer, 1938; 1945; Duncker, 1945), our program of investigation on the cognitive representation of medical knowledge follows the work of Rosch and co-workers

(1973-1977) on categories and prototypes. Rosch views the world as an infinite number of discriminably different stimuli which man, for reasons of limited capabilities, classifies into categories so that nonidentical stimuli can be treated as equivalent. The purpose of category systems is to provide maximum information with the least cognitive effort, referred to by Rosch (1977) as cognitive economy. She argues and empirically supports the idea that the distinctiveness of any category can be increased by the use of prototypes. These prototypical instances (the clearest cases or best examples) contain the attributes most representative of items inside and least representative of items outside the category (Rosch, 1975a, p. 544). Rosch's studies of memory do not contain, nor pretend to build, a theory or a model of memory structures or even less of memory processes, but they do provide a conceptual framework and the methods to acquire empirical evidence capable of constraining an eventual model or theory of cognitive representation.

The Study: This paper will report on a study of the representation of medical knowledge in general practitioners and medical students using categorization and free recall techniques. Preliminary results taken from a group of general practitioners indicate that emphysema, pneumonia and asthma are prototypes of respiratory disorders while aspergillosis and gas embolism are peripheral members of the category, a distinction not so clearly evident to medical students. A second study (in progress) deals with the identification of the source of diagnostic error attributable either to a process error or to a knowledge deficiency. Preliminary results show that most often knowledge is present in the clinician's memory, yet not used at the time of making a diagnosis. Further investigation should clarify whether missed diagnosis is due to poor reasoning habits, such as mental sets (Luchins, 1942), or to the inherent structure of the clinician's store of medical knowledge, such as the inaccessibility of peripheral items in a category.

Medical Problem Solving: A Knowledge-Based Approach

David B. Swanson, Ph.D., School of Medicine, University of Rochester
Paul E. Feltovich, Ph.D., LRDC, University of Pittsburgh
Paul E. Johnson, Ph.D., College of Education, University of Minnesota
James E. Moller, M.D., Medical School, University of Minnesota

Recent research at Michigan State University (Elstein et al., 1978) and McMaster University (Barrows et al., 1977) has clarified the general nature of medical problem solving. Problem solving begins with the elicitation of a small number of cues via questions. These cues suggest a small number of diagnostic hypotheses. Additional cues are elicited and interpreted with respect to the dianostic hypotheses, perhaps suggesting new or revised hypotheses. At some point, hypotheses are evaluated with respect to their relative ability to account for cues, with diagnosis and treatment then determined. This hypothetico-deductive characterization of the problem solving process globally describes expert and novice problem solvers alike. There is large variation across physicians and cases in details of cue elicitation, cue interpretation, hypothesis generation, and hypothesis evaluation processes. The variation is not associated with different levels of physician diagnostic expertise; it appears to be due to problem-specific differences in cases and individual differences in physician knowledge of those problems. The group at McMaster found that actual hypotheses used (as opposed to number used) differentiated successful from unsuccessful problem solvers, with successful problem solvers using hypotheses more suggested by patient findings. The Michigan State group found that "accuracy of interpretation" of data resulted in a similar differentiation. These findings implicate knowledge of medical problems and organization of that knowledge as key elements in successful problem solving.

The research described here uses a similar hypothetico-deductive characterization of problem solving. However, unlike most previous research, which took place in broad domains like internal medicine, this work examines problem solving in the subspecialty of pediatric cardiology. By limiting the scope of the effort, it has been possible to focus on the structure of physician knowledge and the interaction of that structure with case characteristics, the areas implicated as important in problem solving by previous research efforts. Using process-tracing methodology, this knowledge-based approach has resulted in a theory of problem solving which identifies categories and subcategories of clinical knowledge and the role each plays in problem solving (Johnson et al., 1979).

Two major studies using this knowledge-based approach to problem solving have been carried out. The first involved the construction of a computer model which diagnosed cases of congenital heart disease (Swanson, 1978). The various components of the theory were represented as information processes and structures in the model. Twenty patient management problems were diagnosed aloud by a pediatric cardiologist who collaborated on the project. Verbal reports were audiotaped, transcribed, and analyzed to provide a detailed trace of the knowledge and processes used in problem solving. Eleven of the twenty cases were used to develop the idiographic aspects of the model (content of the simulated knowledge base, primarily). Comparison of physician and computer model behavior on these eleven cases revealed perfect agreement on final diagnosis, 85% overlap in the diagnostic hypotheses used by each, and excellent agreement on the findings which caused generation of hypotheses. The remaining nine cases were used in cross-validation of the model. Agreement on the above dependent measures dropped to approximately 70% of the former level, but was substantial, supporting the theory on which the computer model was based. Physician problem solving behavior was demonstrated to be predictable, given a knowledge-based theoretical perspective.

The second study (Feltovich, 1980) investigated the development of diagnostic expertise. Four subjects in each of three groups (medical students who had taken a pediatric cardiology elective, third-year residents in pediatrics, and pediatric cardiologists/third year fellows in pediatric cardiology) diagnosed aloud five patient management problems. The problems were designed to obtain data on knowledge-based differences hypothesized to discriminate the three groups. Each problem had associated with it a logical competitor set (LCS) of potential diagnoses identified through theoretical analyses of the cases. An expert form for diagnosis of each case was developed; it involved full, active use of the LCS as hypotheses during the course of the case. An expert substance for diagnosis was also identified; it defined accuracy of interpretation of key data items for LCS members. Analysis of transcripts of verbal reports from problem solving sessions revealed the hypothesized trends in utilization of expert form and substance. The expert group displayed more frequent use of expert form and substance than the medical student group, with the third group falling in between, again demonstrating the utility of a knowledge-based view of problem solving.

The key elements and strengths of the knowledge-based approach derive from its ties to process-tracing methodology. It is non-experimental, non-statistical, richly descriptive, and, in a sense, naturalistic. It directly uses clinical data in clinical context to obtain detailed "on-line" reports of problem solving behavior. The focus on organization and use of clinical knowledge results in a rich, detailed understanding of the interrelationships among knowledge, problem solving process, and problem/case characteristics. Analysis of skilled problem solvers yields sophisticated clinical algorithms, potentially useful in improving practice and in defining instructional content. Analysis of the less-skilled provides needed additional information, useful in remediating problem areas in practice and in clinical instruction. Weaknesses of the approach also follow from its process-tracing ties. There is reliance on introspective verbal reports, a potential source of difficulty if cross-validating evidence is not sought. Even assuming the accuracy of the verbal reports, analysis of the associated voluminous transcripts is time consuming and poses interpretation difficulties. The description of problem solving which results from such efforts requires additional justification and analysis to become a prescription for clinical practice and design of instruction. However, these difficulties seem surmountable and well worth the effort, because the approach provides key information not yet available through alternative views of medical problem solving.

REFERENCES

Barrows, H., Feighner, J., Neufeld, V., and Norman, G. An Analysis of the Clinical Methods of Medical Students and Physicians. Report to the Ontario Dept. of Health and Physician's Services Inc. Foundation, 1977.

Elstein, A., Shulman, L., and Sprafka, S. Medical Problem Solving. Cambridge: Harvard University Press, 1978.

Feltovich, P. The Role and Development of Medical Knowledge in Diagnostic Expertise. Ph.D. thesis, University of Minnesota, 1980.

Johnson, P., Feltovich, P., Moller, J., and Swanson, D. Clinical Expertise: Theory and Data from Diagnosis of Congenital Heart Disease. Paper presented at AERA Annual Meeting, 1979.

Swanson, D. Computer Simulation of Expert Problem Solving in Medical Diagnosis. Ph.D. thesis, University of Minnesota, 1978.

SYMPOSIUM

PERSPECTIVES ON THE ROLES OF OFFICES OF MEDICAL EDUCATION IN THE 1980's

Chairmen: Emil Petrusa, Ph.D.
Rose Yunker, Ph.D.

Participants: Frank Schimpfhauser, Ph.D.
Charles Dohner, Ph.D.
T. Joseph Sheehan, Ph.D.
Harold G. Levine, M.P.A.
Stephen Abrahamson, Ph.D.

7

PERSPECTIVES ON THE ROLES OF OFFICES OF MEDICAL EDUCATION IN THE 1980's

Rose Yunker, Ph.D., and Emil Petrusa, Ph.D.

The University of Texas Medical Branch
Galveston, Texas

Although the first office of medical education was established at the University of Illinois more than 20 years ago, the majority of them have been operating for less than 10 years. Despite their relative youth these educational resources have become influential in medical education settings. A wide variety of approaches and areas are presently being addressed by these offices. This symposium will describe the current state of offices of medical education and the future directions four such offices from Connecticut, California, Texas, and Washington will take in the coming decade as they continue their impact on medical education. Thus, the two main objectives of the symposium are:

1. To provide a quantitative summary of the present organization and resource allocation of 30 offices of medical education.

2. To present four perspectives on the roles of these offices in the 1980's.

OFFICES OF RESEARCH IN MEDICAL EDUCATION: PRESENT FORM, FUNCTION, AND FUNDING

Frank Schimpfhauser, Ph.D.
S.U.N.Y. at Buffalo School of Medicine

Background

Twenty years have passed since the first formal Offices of Research in Medical Education were established in U.S. Medical Schools. Begun principally as an "experiment" in medical education, increasing numbers of physicians, educators and medical school administrators have embarked on efforts which to varying degrees have supported and emphasized, (1) educational research activities through the systematic collection of student and institutional data, (2) the assessment and improvement of curriculum through the application of educational evaluation methods and measurement techniques, (3) the development of educational methods programs for purposes of improving the instructional skills of medical school teachers and (4) in some instances, providing assistance in the design and implementation of grant proposals for purposes of developing and monitoring curricular innovation. In a broad sense, the goals of the offices have been to support the overall educational mission of the school (primarily at the undergraduate level) by providing formal or informal assistance and consultation to faculty and faculty committees responsible for the teaching and learning which occurs in the medical school environment.

This presentation provides data reported by Director's of Offices of Medical Education relative to the current status of their units. Caution should be taken in interpreting these data since each units form, function and funding pattern is dependent not only upon administrative placement within each milieu, but upon its goal and mission emphasis. Qualitative assessments derived from these data would be fully inappropriate. The data is only intended to serve as a general comparative index of office form and function.

Method

In an effort to quantify factors relative to form (structure), function (activities) and funding (sources of support), a structured survey instrument was developed and mailed to the Directors of 51 recognized Offices of Research in Medical Education in October 1979. Each Director was requested to respond to the following content items:

1. When the Office was established.

2. Whether there have been any shifts in mission.

3. Office placement within the medical school structure.

4. Office mission, goals and objectives.

5. Process used to convert goals into day-to-day activities.

6. Unit functions as a percent of total unit activity (educational research, program evaluation, instructional development, faculty development, direct teaching, educational services, professional staff development, other).

7. To whom the Director reports.

8. Administrative sub-structure within the unit.

9. Professional and major technical staff FTE and their major activity.

10. Support staff FTE and their major activity. (clerical)

11. Percent of office support derived from state funds, grants/contracts and other.

12. Anticipated future: Staff (stable, growing, declining)
    Funding (stable, growing, declining)
    Problem Areas (recruitment, retention, funding, etc.)

13. Miscellaneous:

- active grants/contracts
- active research and evaluation projects
- unit relationship to larger university
- unit relationship to other Offices of Research in Medical Education
- possible development and expansion activities

Results

A total of 30 office directors responded to the mailed survey questionnaire for a survey return of 59 percent. Establishment dates ranged from the years 1959 to 1978 with the largest number of offices being established over the past decade. The number of professional staff employed, as indicated in Figure I was extremely variable and for interpretative purposes might generally be classified as small (less than 10), medium (11-20), and large (above 21). From supportive data included under "major activity", it was clear that within some units, a number of professional-technical personnel were included, particularly when the Office mission included responsibility for activities such as media production, A-V support, instructional development, and test administration and scoring services. Clearly the largest proportions of time allotted to unit activity were in the areas of educational evaluation and instructional development with sixteen percent each given to educational research, and educational services.

Departmental status was indicated by three offices and was roughly defined as those units sporting a "departmental" designation and providing degree granting programs. Fourteen units were classified as "offices"; seven as "divisions"; three as "centers"; two as "programs"; and one as a "unit". Units indicating growth in terms of staff and funding numbered seven, while eighteen indicated relative stability (based on anticipated level of continued grant support) and five indicated a relative decline. Ten units reported that either regular or informal relationships were enjoyed with other Offices of Research. Twenty indicated little, if any, interinstitutional relationship. Fourteen offices indicated a relationship or dependency on other "professional" educational units, departments, or centers within their own University while six indicated little or no intrauniversity relationship. Funding patterns reported were also extremely variable. The percent of extramural grant support dollars covering direct and indirect costs ranged from zero to 92 percent with the average being 25 percent of the total office budget.

Conclusions

The ability or inability to draw specific conclusions from the data reported is rather obvious. What might be generally concluded, however, is that Offices of Research in Medical Education have presumably grown in number due to an increasingly recognized need for providing educational expertise, whether it be in supporting a research mission, or providing services in the areas of instructional support, improvement, ~~curriculum~~ development, and program evaluation. The stated goals and objectives of the offices appear to be somewhat similar, however the form (structure), function (activity) and funding (support) patterns appear to be quite varied.

FIGURE I

OFFICES OF RESEARCH IN MEDICAL EDUCATION SUMMARY DATA 1979-1980

Year Established	Current Office Designation	Location	FTE[1]	Dept Status	Grants[2]	% of Total Office Activity: ER	EV	ID	FD	T	ES	PD	O
1959	Center Educ. Development	Illinois	65		4	10	2	35	10	10	10	3	25
1963	Div. Res. in Med. Education	California	15	*	75	20	5	10	10	5	40	0	10
1965	Office of Medical Education	California	3		25	25	10	15	0	25	25	0	0
1965	Div. Res. & Eval. Med. Ed.	Ohio	8		15	25	30	15	10	5	10	5	0
1966	Off. Med. Ed. Res. & Eval.	Michigan	20		27	Mix							
1967	Div. Std.'s in Med. Educ.	Canada	3		20	5	20	20	20	15	20	0	0
1967	Div. Ed. Commun.	New York	22		0	10	10	10	5	5	50	5	5
1968	Div. Ed. Serv. & Res.	Michigan	15		30	20	20	20	10	20	10	0	0
1968	Dept. Res. In Health Educ.	Connecticut	12	*	47	25	20	20	10	20	0	3	2
1968	Off. Res. In Med. Education	Washington	15		92	15	30	30	10	5	5	5	0
1970	Div. Res. In Med. Education	Pennsylvania	2		5	5	45	20	0	0	30	0	0
1970	Off. Educ. Studies	North Carolina	8		48	10	15	15	15	15	10	15	5
1971	Off. Res. in Med. Education	Texas	6		20	Mix							
1972	Off. Med. Education	Maryland	25		25	Mix							
1972	Off. Med. Education	Arizona	8		0	20	25	20	20	0	10	5	0
1972	Prog. Educ. Development	Canada	5		40	40	20	20	10	10	0	0	0
1972	Off. Ed. Resources & Res.	Michigan	70		31	20	20	25	5	5	15	10	0
1973	Ed. Planning & Dev. Prog.	Virginia	10		36	10	15	15	25	2	10	3	20
1973	Div. Ed. Support & Dev.	Canada	2		5	10	15	10	10	15	15	5	20
1974	Office of Medical Education	Oregon	1		0	Mix							
1974	Office of Medical Education	Illinois	2		15	15	15	20	20	10	20	0	0
1974	Dept. P.G. Med. & CME	Ohio	8	*	10	10	20	5	20	15	20	5	5
1974	Center for Med. Education	Canada	5		0	25	25	0	5	0	20	5	20
1975	Ed. Development Unit	Illinois	11		65	10	10	15	30	20	15	0	0
1975	Center Ed. Dev. & Eval.	North Carolina	4		0	5	5	75	5	2	3	5	0
1975	Off. Educ. Development	Alabama	15		40	Mix							
1975	No formal unit name	Missouri	6		0	Mix							
1976	Off. Med. Ed. Res. & Eval Unit	New York	5		30	20	40	10	10	5	5	5	5
1978	Off. Ed. Res. & Serv.	North Carolina	3		33	13	15	20	10	10	30	2	0
1973	Off. Ed. Res. Eval. & Dev.	Texas	4		0	5	25	40	5	2	5	0	12
MEAN			12		25	16	19	20	11	10	16	3	5
MODE 1972			15		0	10	20	20	10	10	10	5	5

[1]Full time equivalent professional staff
[2]Percent office support from grants/contracts
ER - Educational Research
EV - Educational Evaluation
ID - Instructional Development
FD - Faculty Development
T - Direct Teaching
ES - Educational Services
PD - Professional Staff Development
O - Other

PERSPECTIVE FROM UNIVERSITY OF WASHINGTON

Charles Dohner, Ph.D.
Office of Research in Medical Education
Seattle, Washington

Questions concerning the future of research in medical education during the coming decade are not easy to answer. The best one can hope to do is identify the chief factors that are beginning to affect academic institutions of medicine and then hypothesize about the extent of their continuing influence. Two factors which have already begun to make themselves felt and are most likely to continue to do so in the coming years are: the issue of funding and the issue of manpower needs for academic positions within medical education settings.

Concerning funding, the priorities of federal and foundation granting agencies have had and will continue to have considerable influence on research and evaluation projects undertaken within the tertiary health care center. The present thrust of these agencies is increasingly problem-centered and content-specific. Proposals which use interdisciplinary approaches to study specific health issues such as gerontology, alcoholism, and nutrition are attracting funds. Patient education, compliance, and preventive health care are subjects of growing interest. The decade of the eighties is most likely to see an increase of research and evaluation projects focused on specific issues of a similar type.

Implicit in the foregoing hypothesis is a corollary, that of decreasing funding and hence decreasing research and evaluation focused on innovations in medical education, particularly at the undergraduate level. Our office has experienced a growing conservatism among faculty members with regard to teaching methodology. They are less interested now than in the past in experimenting with new approaches to their undergraduate courses. A shift in emphasis from the training of undergraduates to the training of residents is visible. This, too, can be linked to federal and foundation funding priorities.

The second chief factor influencing our immediate future is the issue of manpower needs in the medical education setting. Two relatively new additions to academic medicine, family medicine, and primary care, are seeking turf on which to build their credibility. Credibility in academic medicine has traditionally been judged by research activity. Both family medicine and primary care generalists within more established departments are looking for faculty members who can initiate research projects from the scope of their ambulatory clinic encounters. Some of these positions will be filled by the doctors now being trained as clinical scholars. Others, as a result of the funding constraints being felt everywhere, will be filled by in-house candidates who will need on-the-job training in doing this kind of research. Education, collaboration, and evaluation in these research projects will be a priority in the coming decade.

As I indicated when I began these remarks, questions concerning the future of research in medical education do not yield to easy answers. The trends I have identified seem to offer some clues to what the future may look like during the 1980's. I feel confident that the educational expertise necessary to meet these new challenges exists in offices of research in medical education and will be increasingly in demand as we move into the future.

PERSPECTIVE FROM UNIVERSITY OF CONNECTICUT

T. Joseph Sheehan, Ph.D.
Research in Health Education, University of Connecticut Health Center

As we look into the crystal ball at the next decade to determine the shape of our medical education units (offices, divisions, departments), we might first speculate about the general setting in which we work--the medical schools. There seem to be at least three major sources of influence that are affecting the future.

First, it appears to be the view of the policy makers that the manpower problem has been solved, or at least redefined. The number of doctors, it is argued, appears to be adequate; the geographic and specialty distribution remains a problem.

Second, competition for research funding at both the clinical and basic science levels is becoming tougher and tougher. It is not uncommon today to learn that even senior investigators are having trouble with funding.

A third source of influence, in addition to manpower and research, is a growing awareness of the limits of the strictly medical intervention model as the answer to the health problems of this country and a beginning realization that health education, health promotion, and preventive medicine have been badly neglected.

The impact of these influences, in my opinion, will be to increase the amount of time our clinical faculty spends doing direct clinical care while decreasing their involvement in both research and teaching. In the basic science departments, I see an increase in the time devoted to research in those schools where the faculty continues to be competitive for grant support and a corresponding decrease in time available for instruction and the instructional program. In a number of schools, there may even be a thinning of the ranks of basic science faculty as the school is faced with the salary burdens of those who fail to compete successfully for grants. Among the survivors in those schools I would predict heavy teaching loads as they take up the slack left by their departed brethren.

This all sounds rather gloomy, and I hope that my sketch is overdrawn, but if it is accurate, where does this leave our units? First of all, many of the units have a tradition of being more reactive than proactive. They have built their success on responding to the educational support needs of faculty at their institutions. In the early years, it was a matter of creating systems for writing, scoring, and interpreting objective examinations. In many units, as the faculty became more adept and the system operational, the entire function was moved from the unit to the Dean's Office. Lately there has been an increasing call for help in program evaluation, especially the instructional aspects of the categorical disease center grants.

I suspect there is going to be less call for services which directly support the educational program, especially the undergraduate medical education program, and more for collaboration on grant supported evaluation projects. Because of decreasing demands for our services due to increased demands for patient care and research placed on the faculty, our units must become more proactive. Just because we are not asked, does not mean that the needs have disappeared. We will have to become much better at identifying problems, judging their tractability, and obtaining the cooperation of faculty and other support required to solve the problems. What I am saying, quite frankly, is that our units, like other departments, will be expected to obtain more and more of their own support, and in so doing will face several pitfalls.

One very real danger is to lose the sense of mission in favor of the pursuit of survival. By this I mean the ideals of the educational enterprise in which we participate can be subverted by our self-preservation instincts. Education would be much better off if those who have lost the ideals of the quest would look to a change of profession rather than continue as non-believers--apathetic deadwood. There must be an effort to balance our research and other independently supported activities with a renewed sense of service to the educational mission of the institution.

To achieve greater autonomy, however, I do foresee more initiative from our units, with the net effect being a large increment in the percentage of effort devoted to sponsored research projects. Further, I would predict greater numbers of principal investigators coming from our units, with collaborators coming from other departments, a change from current practice. As a corollary, I see research as the primary purpose of many of these efforts, rather than as a spin-off from an evaluative effort by a creative collaborator. An increase, even a substantial increase, in the time and effort we devote to research is warranted. This portends a danger or pitfall for our units that we now are beginning to see for the rest of the school. The schools are in danger of succumbing to the same kinds of problems identified by Flexner years ago. He argued that it was not enough for our students to be taught by clinicians who were part-time teachers. He felt there should be full-time clinical faculty who taught and did research as their major obligation and not just as a sideline. I fear that we are not only reverting to that situation with the new pressures on clinical faculty, but their basic science counterparts may present the other side of the coin: full-time basic scientists who also happen to teach.

PERSPECTIVE FROM UNIVERSITY OF TEXAS MEDICAL BRANCH

Harold G. Levine, M.P.A.
Office of Research in Medical Education
The University of Texas Medical Branch, Galveston, Texas

In some medical schools, ORME's have flourished and expanded. In others, offices have been treated as frills that have been cut back when external funding has run out or the institution has undergone some financial stress. Still others, which include most of the private, research-oriented institutions, the offices have never been created. It may be presumptuous to predict, in the face of such a past history, that eventually educational research and development activity must be an integral part of the basic activities at every medical school.

Part of the problem that educational R&D in academic health science centers has suffered is that many of its concerns are not central to the mission and function of the organization. Some common activities of such units include such activities as faculty development workshops, the development of instructional materials, the improvement of objective tests, the gathering and distribution of student evaluation data, analysis of clinical problem-solving exercises, etc. All of these are worthy activities, but they often function in ways which are peripheral to the central problems of medical education which have only begun to be addressed.

The most serious problem of medical education is that those who serve as faculty of medical schools are socialized by the post-doctoral experiences to be either interested primarily in careers in biomedical research or careers in subspecialty practice or research. This comment is not meant to disparage either of these excellent careers. Biomedical research may ultimately be much more significant to the health of future populations than the training of physicians. However, one cannot be done at the expense of the other. It doesn't do much good to have an excellent center for neonatal care if physicians or nurses do not refer newborns at risk to the center. Most of the activities of ORME's have accepted this faculty preoccupation as a fact of life and have not directly addressed ways of altering faculty preoccupations and providing alternative career pathways in medical education.

Some of the ways I believe that educational R&D centers and personnel will collaborate with physician and biomedical faculty in the future to provide a more balanced functioning of academic health science centers include the following:

1. Serving as faculty in clinical departments. Most clinical departments have assumed enormous teaching responsibilities. Physician faculty need colleagues who are experts in behavioral science, educational psychology, health services research, etc., to perform such key activities as assisting in the evaluation of clinical performance, organizing instructional activities, teaching behavioral sciences, etc.

2. Addressing issues of organizational development and change. Since much of the faculty activity is controlled by subtle norms of faculty socializations and perceptions, much work needs to be done in understanding these norms and dealing with them in ways which will facilitate the institution's ability to respond to its mission.

3. Addressing the institution's responsibilities for continuing education. CME has been treated as an adjunct of the undergraduate and graduate periods of medical education. In the future it will have to be regarded as part of the central mission of the institution. Physicians may well be required to enroll in CME programs affiliated with institutions.

4. Conducting activities in health services research, with especial emphasis on the linkages between method of educating physicians and their ultimate impact on patient care outcomes. We have been evaluating physicians mainly on what they remember from their classes in medical school and not how their training affects the care of patients. In the long run, such an evaluation system can blind us to the unanticipated consequences of certain educational practices. For example, some believe that present educational practice encourages rote learning and discourages attention to issues of preventive medicine, continuity of care, and care of the chronically ill.

Many of the activities we pursue today are intended to eventually impinge on these four issues. However, they often are viewed as ends in themselves, and their relevance to the delivery of health care is lost. I believe that eventually educational R&D will become a vigorous component of academic health science centers.

RESEARCH IN MEDICAL EDUCATION IN THE 1980'S

Stephen Abrahamson, Ph.D.
Director, Divison of Research in Medical Education
University of Southern California School of Medicine

In an effort to predict what may be ahead in the 1980's, this paper first examines the development of such activities of the Division of Research in Medical Education at the University of Southern California School of Medicine from its inception in 1963 to its present status.

DRME - PAST

The "history" of the Division of Research in Medical Education can be roughly viewed in five major periods.

(1963-1965)

These earliest years were characterized by individual consultation on matters related to teaching and learning in undergraduate medical education. These consultations included advice with regard to course planning, item analysis on examinations, and development of checklists and grading scales. In some instances, this consultation involved working with a department or standing committee, not just an individual faculty member.

(1965-1967)

The character of the Division changed significantly with the awarding of three grants: (1) to develop new teaching approaches in neurology (i.e., the "programmed patient" and single-concept, self-instructional films); (2) for the development of a computer-controlled, patient simulator for training anesthesiology residents in endotracheal intubation (later known as "Sim One"); (3) to study a major change in grading and examination practices in the School of Medicine.

(1967-1971)

There then ensued an incredible period of growth, including establishment of a training program to prepare those with a doctorate in education or psychology for careers in educational evaluation in medical education settings; a study to help develop better techniques for certification of pediatric cardiologists; a wide variety of evaluation studies in continuing medical education; establishment of a media service for the School of Medicine; and staff support of a curriculum change in the School of Medicine.

(1971-1975)

During this period of time, the Division found itself more and more involved in health manpower studies, essentially examining practices of physicians. This was also a period of intensifying efforts to systematize services provided internally to the faculty of the School of Medicine (e.g., a "testing service" with an item pool which by 1975 had reached 8000 items).

(1975-1979)

With the concentration on manpower studies and internal services, the activities of the Division seemed to become quite stable and rather static. It is too soon to tell whether this "stability" presages a general decline in activity or not.

DRME - PRESENT

Essentially there are six major areas of activity at the present time. (1) Faculty development programs seem to occupy a major segment of time - both within the School of Medicine and outside. (2) The media services have become stable and involve a major effort in instructional design. (3) Testing services continue to occupy a major portion of time and effort with an item pool of more than 10,000 items. (4) A new area of services has been growing: maintenance of data banks on students and faculty. (5) Programs attempting to provide needed services to minority students are administered by this Division. (6) There is a new effort in simulation, leading toward the development of Sim Two.

DRME - FUTURE

What can be learned from the past? (1) It seems quite clear that federal funding patterns do have an impact on local activities. Sutton's Law obviously affects those who conduct research in medical education! Therefore, it does seem safe to predict that "hard times" are upon us: there will be less money; therefore, there will be less activity. Even during the greatest days of federal support, more lip-service was given to the support of evaluation activities than was made available. Now, it will be worse.

(2) It is quite clear that internal support will be continued in areas of needed services. If a faculty appreciates services in support of its educational program, the administration will undoubtedly provide support for those services. If the administration appreciates certain services, support will be made available.

(3) Despite the disclaimer above that evaluation is more talked about than conducted, serious and comprehensive evaluation studies will be a major part of the future. This is particularly important when one considers the need for sound evaluative data to justify expenditures.

(4) Selected educational innovation will be supported - if it is accompanied by rigorous testing. Obviously, again, the major import here is in the area of economies.

Thus, the future looks bright in some ways and threatening in others. To the extent to which research in medical education can be responsive to institutional needs, it will continue to develop and thrive. To the extent to which research in medical education attempts to "do its own thing" it will meet with possible resistance and rejection - if not extinction.

SYMPOSIUM

A MULTI-INSTITUTIONAL RESEARCH STUDY ON THE USE OF SIMULATION FOR TEACHING AND EVALUATING PATIENT EXAMINATION SKILLS

Chairman: Abdul W. Sajid, Ed.D.

Participants: Michael S. Gordon, M.D.
Joan W. Mayer, M.D.
Gordon A. Ewy, M.D.
Alan D. Forker, M.D.
Abdul W. Sajid, Ed.D.
Joel M. Felner, M.D.
Dorthea Juul
Robert A. Waugh, M.D.

Discussants: Howard S. Barrows, M.D.
Arthur S. Elstein, Ph.D.

A MULTI-INSTITUTIONAL RESEARCH STUDY ON THE USE OF SIMULATION FOR TEACHING AND EVALUATING PATIENT EXAMINATION SKILLS*

Abdul W. Sajid, Ed.D.
Center for Educational Development
University of Illinois Medical Center

The use of simulation for teaching and evaluation has been a major topic among medical educators, and a number of innovative devices and methodologies have emerged. The reactions of medical teachers to these innovations have ranged from total skepticism (as frills, toys) to acceptance as a panacea for all the ills of clinical teaching. Unfortunately, these opinions are based on sketchy data and extremely limited research. The major objective of this symposium is to highlight the findings of a comprehensive, multi-center evaluation study of medical simulation utilizing a cardiac patient simulator (CPS). The simulator was developed by Dr. Michael S. Gordon and his colleagues at the University of Miami School of Medicine.

Specifically, the presentations as well as the discussion will address the following questions:

1) Do students taught with the simulator exhibit equivalent cognitive knowledge to students in traditional cardiology clerkships?

2) Are the clinical skills learned on the simulator transferable to real patients?

3) Are there any differences in patient perceptions of students taught on the CPS and students in traditional cardiology clerkships?

4) Is instruction utilizing the CPS more efficient than conventional patient dependent instruction?

5) How do other simulation techniques compare to the CPS (problem boxes, Sim One, computer, etc.) in the teaching and evaluation of bedside clinical skills?

6) What are the implications for using simulation devices in clerkships where students have ready access to real patients?

*This study was funded by the National Institutes of Health contract number NIH-N01-HN-62969.

OVERVIEW OF THE CARDIAC PATIENT SIMULATOR STUDY

Michael S. Gordon, M.D., University of Miami School of Medicine
Gordon A. Ewy, M.D., University of Arizona Health Sciences Center
Alan D. Forker, M.D., University of Nebraska Medical Center
Joan W. Mayer, M.D., University of Miami School of Medicine

Sound clinical judgment involves the application not only of broad knowledge and understanding, but also of complex cognitive and psychomotor skills and relevant professional habits and attitudes. Conventional methods of instructing and evaluating students, residents, and clinicians are often less than optimally suited for developing these competencies. Analogous dilemmas in other fields have sometimes been resolved by the introduction of sophisticated simulators. These include business and management games for business executives, war games and training exercises for military personnel, and flight simulators for pilots and astronauts. Although simulation does not duplicate life, it does place the individual in a life-like situation that requires his/her participation in initiating and carrying through a sequence of inquiries, decisions, and actions.

Simulation techniques for teaching and testing have also been developed in medical education. These include paper and pencil simulations (patient management problems, problem boxes), reproduction of visual and auditory stimuli (breath, heart, and abdominal sounds, photographic reproductions, videotapes), three dimensional models (e.g., pelvic and eye models), automated mannequins, computer-aided simulations, and live simulations. These techniques avoid such drawbacks as the unavailability of patients for display of specific diseases at specific moments in the curriculum schedule, the inconvenience and risk to patients, the stress to students, and the expense and complexity of using patients.

The literature concerning the variety of simulators is increasing rapidly, yet little empirical evidence of the effectiveness and validity of simulation has appeared. Most of the studies are based on small samples and do not test the application of skills learned on simulators to real patient situations.

The current study was undertaken to comprehensively evaluate a cardiology patient simulator (CPS) developed under the direction of Dr. Michael S. Gordon at the University of Miami School of Medicine. The CPS is a sophisticated, life-size mannequin that simulates the bedside encounter with a patient. Physical findings simulated include blood pressure, venous and arterial pulsations, precordial pulse abnormalities, and the auscultatory events associated with each disease state. Ancillary data such as history, laboratory results, medical and surgical therapy, and the pathology and epidemiology of each disease are presented on slide programs via a self-assessment format. Currently, the mannequin is programmed for twenty major cardiac problems.

The mannequin, therefore, presents the student with the opportunity to obtain data about a patient, perform a cardiac examination, make judgments about therapy and diagnosis, and receive feedback on various aspects of his/her performance.

Cardiologists from the medical schools of the universities of Arizona, Duke, Emory, Miami, and Nebraska participated in this study along with educators from the Center for Educational Development of the University of Illinois Medical Center. The main objectives of the research study were to:

1) Assess the cognitive knowledge of CPS and non-CPS students

2) Assess the cardiac examination skills of CPS and non-CPS students

3) Assess student and faculty attitudes towards the CPS

4) Determine the efficiency of instruction with the CPS

5) Determine the technical reliability of the CPS.

Subjects were fourth year medical students enrolled in cardiology electives at the five sites. Clerkship groups were kept intact with approximately half the clerkships at each site utilizing the CPS and half not. When students registered for their clerkship, they did not know if they were or were not going to be taught with the CPS. Time was available for students in non-CPS clerkships to use the simulator after post-testing. Clerkship length varied by site. Three sites had four week programs, one had a six week program, and one had an eight week program.

All students were pre-tested on a 90 question multiple choice test and one patient management problem and performed a cardiac examination on two randomly selected CPS patients. They were post-tested on a parallel multiple choice test and patient management problem and examined two live patients in addition to two CPS patients. In addition, students and faculty were asked to fill out an attitudinal questionnaire. A two-way analysis of variance (treatment X site) will be used to analyze the multiple choice, patient management problem, and skills test data.

Efficiency of instruction is based on time data. Students and faculty were asked to record how much time was spent on various activities for both group and individual CPS and patient sessions. These logs were kept for one week of each clerkship. Data on the technical performance of the simulator was collected during each CPS clerkship.

Data collection began in July 1979 and was completed in June 1980. The total sample size for all five sites was about 300 medical students and 15 faculty. Each clerkship involved from five to ten students. There were no major problems with data collection, and there was very little attrition of subjects between pre- and post-testing.

In summary, this study addressed the question of whether medical students can acquire cognitive knowledge and clinical skills through the use of a simulator in an effective and efficient manner when compared to students in conventional cardiology clerkships. More critically, the study provided data on the issue of transference of clinical skills learned on a simulator to real patients.

351

USE OF THE CARDIAC PATIENT SIMULATOR FOR TEACHING

Abdul W. Sajid, Ed.D., Center for Educational Development, University of Illinois Medical Center
Joel M. Felner, M.D., Emory University School of Medicine

A common core of objectives on which all collaborating institutions agreed was developed. These objectives reflected the skills and cognitive knowledge commonly addressed in fourth year medical school cardiology clerkships. Beyond this common core, additional objectives and methods of instruction varied among institutions. However, within an institution the objectives were the same for CPS and non-CPS clerkships. CPS instruction was substituted for other scheduled activities (mainly patient contact activities) and not added on to the regular teaching schedule so that the experimental and control groups had equivalent instructional opportunities. It should be noted that the CPS students did have patient contact through bedside teaching rounds and individual work-up of patients.

In order to gain information on how the CPS was actually used for instruction, students and faculty were asked to record how much time they spent on various activities during one week. In CPS clerkships, students reported on group and individual CPS and patient sessions and faculty on group CPS and patient sessions. For non-CPS clerkships students filled out group and individual patient logs and faculty filled out group patient logs. Some of the main categories of behavior included:

<u>Group Patient Logs (Student and Faculty)</u>
- Total time scheduled for rounds
- Time at bedside examining patient
- Reviewing/discussing history, physical, laboratory, diagnosis, and management
- Faculty demonstrates/reviews examination skills
- Availability of patient data, e.g., chart, laboratory results
- Reason for selecting patient

<u>Group CPS Logs (Student and Faculty)</u>
- Total time scheduled for CPS instruction
- Time spent examining CPS
- Reviewing/discussing history, physical, laboratory, diagnosis and management
- Faculty demonstrates/reviews examination skills
- Technical problems with CPS

Individual Patient Log (Student)

- Time spent finding patients
- Time spent doing work-up
- Problems in working up patients

Individual CPS Log (Student)

- Time spent working with CPS
- Time spent learning to use, setting up CPS
- Technical problems with CPS

A profile of how time was spent in CPS versus patient clerkships will be developed. Two key questions are (1) does the CPS increase the amount of productive learning time because it eliminates such problems as unavailable and uncooperative patients and missing data; and (2) does the CPS increase the amount of time students spend in "hands on" activity because of the availability and "cooperativeness" of the mannequin. Congruence between faculty and student observations will also be examined as well as site differences.

The effectiveness of the CPS for teaching cognitive information will be determined on the basis of performance on a 90 item multiple choice test and a patient management problem. The test instruments were developed to assess attainment of the core clerkship objectives.

Further evidence of the mannequin's effectiveness will be provided by attitudinal data collected from faculty and students. They were asked to respond strongly agree, agree, disagree, strongly disagree, or no opinion to 19 statements about different aspects of the CPS. There were also some open-ended questions, and faculty were asked to write a statement about their teaching experiences with the simulator.

In terms of technical reliability, time spent on technical checkout, problems encountered, solutions applied, and time allotted for correcting the problem as well as user problems were recorded at each site.

In summary, this presentation will discuss how the CPS was actually used for both group and individual instruction based on faculty and student log data. Comparisons with non-CPS instructional activities will be made. Evidence on the effectiveness of the mannequin based on cognitive performance and attitudinal information and on its technical reliability will also be presented.

USE OF THE CARDIAC PATIENT SIMULATOR FOR EVALUATION

Dorthea Juul, Center for Educational Development, University of Illinois Medical Center
Robert A. Waugh, M.D., Duke University Medical Center

The importance of evaluating clinical skills is generally recognized yet the assessment of "hands on" performance is often neglected because of the difficulties it poses. It is often a problem to obtain appropriate patients who are willing to submit to multiple examinations. Faculty may be hesitant to conduct skills evaluation because of the time required to observe examinees. There are also the standard testing concerns of validity, reliability, and standardization.

For these reasons, various simulation approaches have been developed such as paper-and-pencil patient management problems and actors trained to present with certain conditions. Another profitable approach for testing skills may be the use of devices such as the CPS.

Simulation has several advantages. First, it may enhance validity by placing the examinee in a situation more closely resembling medical practice than conventional testing formats such as multiple choice questions. More importantly, simulation makes it possible to precisely select the examination context. Testing is no longer dependent on the availability and good will of patients.

Simulation may improve reliability by eliminating irrelevant factors that are often present in real life, e.g., interruptions by other staff, uncommunicative patients, and by focussing on the elements of primary concern. Preselection of the tasks may also improve reliability by allowing for development of specific performance criteria and training of examinees.

Simulation allows for standardization of the test situation. All examinees have to cope with the same problems without patients being subjected to repeated examination.

Although simulation is not appropriate for measuring all aspects of medical student/physician performance, it does seem to have several advantages, particularly for skills assessment. Yet an important question remains. How does performance on the simulator correspond with performance on real patients?

To determine if simulator performance is a valid measure of performance on real patients, students in this study performed a cardiac examination on both the CPS and real patients. They examined two CPS patients at the beginning of the cardiology clerkship and two CPS diseases and two real patients at the end. The CPS diseases were randomly assigned.

It was decided after much deliberation that faculty would not observe the students. Rather, a form was developed by the cardiologists participating in the study to cover all significant aspects of the clerkship objectives for the cardiac examination. The major categories were venous pulsations, respiratory rate, pulse rate and rhythm, blood pressure, carotid and femoral pulses, precordial movements, apex impulses, heart sounds, murmurs, and diagnosis. Students recorded their findings on the form, and points were awarded for correct answers on the basis of the relative importance of each item.

Master keys were developed and computerized for the twenty CPS diseases. Keys for the patient examinations were prepared by faculty on the test day. To check the reliability of these keys, an independent examiner was asked to examine a patient and complete a key. Preliminary analysis demonstrated that the percentage of interrater agreement tended to be fairly high ($>.90$).

Correlations between scores on the CPS and patient examinations will be presented, as well as other test data. In addition, patients were asked to evaluate their student examiners on several factors, and this information will be presented. Correlations between test scores and patient satisfaction will be reported, and comparisons will be made between students in CPS and non-CPS clerkships on the basis of test performance and patient satisfaction.

In summary, this presentation will focus on the use of the CPS for the evaluation of cardiac examination skills, and performance on the simulator will be compared to performance on patients. General recommendations on the CPS's role in evaluation will also be made.

355

SYMPOSIUM

RECRUITMENT AND SELECTION OF PHYSICIANS FOR PRIMARY CARE AND RURAL PRACTICE -- RESULTS FROM THE PROJECT TALENT LONGITUDINAL STUDY

Chairman: Sandra R. Wilson, Ph.D.

Participants: Sandra R. Wilson, Ph.D.
Lauress L. Wise, II, Ph.D.
Frances B. Stancavage

Discussants: Lee Sechrest, Ph.D.
Paul Elliot, Ph.D.

RECRUITMENT AND SELECTION OF PHYSICIANS FOR PRIMARY CARE AND RURAL PRACTICE -- RESULTS FROM THE PROJECT TALENT LONGITUDINAL STUDY

Organizer: Sandra R. Wilson, Ph.D.
Director
Health Research Group
American Institutes for Research
Palo Alto, CA

Since 1960, the American Institutes for Research have conducted a longitudinal study of a nationally representative sample of 400,000 high school students known as Project TALENT. From the TALENT grade cohort, approximately 1670 physicians and nearly 3000 medical school applicants emerged who passed through the admissions process and entered medical school (if they did) in the period between about 1964 and 1968. The data available on these individuals are extensive and include abilities, interests and activities, background, educational and occupational plans, and other characteristics in high school, as well as follow-up data covering educational activities and other pursuits at a number of intervals post-high school.

In the present study these data have been linked to AAMC and AMA records that provide information on characteristics at application to medical school, the school attended, post-graduate training and present practice. This linkage permitted investigation of the early characteristics of physicians now in various specialties and practice settings, and the interrelationship of personal characteristics and educational experiences to career outcomes. Because of the availability of the samples of applicants (including non-entrants) and of high school students in general, it was also possible to compare high ability high school students, applicants, entrants and physicians in terms of the proportions in each group who were similar to primary care and/or rural physicians.

The fact that the physician sample is nationally representative, that the data were gathered longitudinally beginning back in high school, and that there is available for comparative analyses a representative sample of both applicants and of the population as a whole, including non-physicians, makes this study unique as a basis for studying medical career development.

The purpose of this symposium is to present results of this study for the first time and to relate the findings to current issues in the medical school admissions process and to efforts to alter the geographic and specialty distribution of physicians. The specific objectives of the symposium are:

1. to describe the sample, data base, and the general background of the study,
2. to describe the selection and refinement of the outcome measures,
3. to describe the general methodological approach to data analysis and the specific analytic techniques employed,

4. to describe the specialty and location outcome measures used in the analyses,
5. to describe the results most relevant to specialty and location choice and issues in medical manpower development,
6. to discuss the implications of the findings for recruitment, selection and training of physicians, and
7. to briefly discuss plans for further research using this sample and data base.

The participants in the symposium will include the organizer, who directed the research program, and the principal members of the research staff, Dr. Lauress L. Wise and Ms. Frances B. Stancavage. The discussants will be Dr. Lee Sechrest, who will comment on methodological aspects of the study in the context of other longitudinal studies of career development in medicine and other professions, and Dr. Paul Elliott, who will consider the implications of the findings for efforts to recruit, select and train physicians to fulfill particular manpower needs. Dr. Wilson will give a brief introduction and background of the study. Dr. Wise will describe the TALENT sample and subsamples used in this study, the data bases, the practice outcomes and analysis techniques. Ms. Stancavage will describe the differences among practitioners in various specialties and locations, and Dr. Wilson will discuss the findings concerning the overall selection process from high school through medical training and their implications.

Study of the Project TALENT Physicians -- Samples, Data Base, and Analysis Approach

Lauress L. Wise, Ph.D., Senior Research Scientist and Director, Project TALENT, American Institutes for Research, Palo Alto, CA.

The study utilized a large body of existing data on a nationally representative sample of approximately 3000 medical school applicants including 1601 physicians who were identified among the more than 400,000 individuals included in the Project TALENT Longitudinal Study of American High School Students. The Project TALENT cohort, which is a national probability sample, was first selected and studied intensively in 1960 when these students were in high school. Data were collected on the interests, abilities, activities, background, and plans of all of the students in the course of two full days of testing and questionnaire completion. The cohort was also surveyed at three points subsequent to 1960 to determine the nature of their post-high school educational, occupational, and personal experiences.

Physicians in the TALENT cohort generally received their medical degrees between 1968 and 1972, and all but about 200 (who were still completing residencies around 1978) were found to be in practice or other medical employment. A variety of resources, including data from Project TALENT, the Association of American Medical Colleges, and the American Medical Association, were used to identify those members of the TALENT cohort who had been medical school applicants, those who entered medical school and those who became physicians, and to classify this last group by specialty and practice location types. Data from each of these three sources (Project TALENT, the AAMC, and the AMA) were compiled on each person identified, and additional data on the characteristics of their colleges, medical schools, and current practice counties were added to the individual records.

In the first phase of the study, investigation centered on the physician sample. Particular outcome groups of interest were identified among the physicians using data from the AMA Physician Masterfile. The dimensions on which the outcome groups were defined included specialty, level of specialization, geographic region of practice, urban-rural character of the practice county, and county physician-population ratio. Several hundred variables representing characteristics of the individual, and others representing characteristics of the college and medical school the individual attended, were compared among the outcome groups. Variables that were significantly related to these outcome dimensions were thus identified, and profiles of the unique characteristics of physicians in different specialties, levels of specialization, and communities of differing sizes were prepared using these variables. The profiles, available in the complete Final Report, provide a comprehensive description of the personal attributes, backgrounds, educational institutions, and location histories that characterize different types of physicians. Selected profiles will be discussed in the second presentation in this symposium.

The variables included in the profiles were also entered into a series of discriminant analyses to determine the extent to which physicians in different specialties and locations could be reliably discriminated from each other and the accuracy with which their specialty or location could be predicted using discriminant functions containing information that would have been available either during high school, as of the end of college, or as of the end of medical school. The discriminant analysis techniques will be described in the presentation in sufficient detail that their use in the second phase of the study can be understood.

The second phase of the study utilized both the sample of applicants to medical school and a specially selected stratified sample of approximately 8000 Project TALENT participants. Participants selected for this latter "manpower" sample were all respondents to the 11-year post-high school TALENT follow-up, so that information was available on their post-secondary education and present occupations. They were selected in such a way as to constitute a probability sample of all 1960 high school students. Analyses of this sample permitted the derivation of estimates of the numbers of individuals--among the medical school applicants, and in the cohort as a whole--who had characteristics similar to those students of who became general/family physicians and rural practitioners. The findings will be described in the third paper in this symposium.

Profiles of Medical Practitioners and Prediction of Specialty and Location

Frances B. Stancavage, Senior Research Associate, American Institutes for Research, Palo Alto, CA

The first major finding of the study was that physicians in different specialties, at different levels of specialization, and in different sized communities were discernibly different in terms of their personal and sociodemographic characteristics during high school. These significant differences persisted through college and medical school and were apparent, both in the characteristics of the physicians and characteristics of the institutions in which they were educated. The types of differences found between those in different specialties and levels of specialization include:

High School:	Occupational and activity interests in high school
	Academic abilities and performance in high school
	Academic emphasis of curriculum in high school
	Sex, religion, and size of family of origin
	Geographic region and community size background
	Leadership experience and certain personality characteristics
	Post-high school educational and occupational aspirations
College:	Type of college attended (control, selectivity)
	Geographic location of college and proximity to high school community
	Sources of financial support during college
	Continuity of attendance and age at entrance and graduation
	College major
	Grades and MCAT test scores
Medical Training:	Type of medical school (control, selectivity, AAMC cluster)
	Geographic location
	Location of graduate medical education

Physicians practicing in different-sized communities differed on the above dimensions. In addition, they differed in the occupations and educational levels of their fathers. These results indicate that a wide variety of early individual differences play a substantial role in both specialty and location choice and that many of the same characteristics are related to these two choices. The characteristics of two groups of practitioners of special interest are summarized below; actual profiles of these two groups and other specialties will be displayed and discussed in the presentation.

General and family practitioners, who comprised only about 7% of the physicians in the cohort, were the most readily differentiated specialty group. While near the physician average in overall general academic ability in high school, they clearly had less extensive academic preparation and somewhat lower achievement in mathematics. They came from smaller communities and from the Southeast, Plains, Far West, Southwest, and Rocky Mountains more often than other physicians. GP/FPs also came from somewhat lower SES backgrounds, from Protestant or Catholic but not as frequently from Jewish backgrounds, and from large families more frequently than did other physicians. The proportion of females in this specialty was relatively low compared with the proportion in medicine as a whole.

The pattern of interest and information test scores of GP/FPs is also unique and perhaps related to both their background and predominant sex. Their interests in high school were very low in the area of public service (governmental occupations), somewhat lower than those of other physicians in the biological sciences (including medicine), art and literature, and relatively high in mechanical work, skilled trades, labor, and farming. They knew less about math and the arts, but much more about the Bible, farming, hunting and fishing than other physicians-to-be. Their educational expectations were lower when they were in high school, their occupational expectations more focused on non-medical and even non-science fields, and their expected college majors were more frequently in math, the physical sciences or in social studies/humanities than those of other physicians.

Future GP/FPs were as likely to have enrolled in public colleges as other physicians, but those attending private institutions were more likely to have attended religious-affiliated (denominational) colleges. They were relatively less likely to have attended the most selective colleges and universities, somewhat less likely to have majored in chemistry, and they achieved somewhat lower MCAT scores than other physicians. The colleges they attended were less frequently located in the Northeast and were more often in the region in which they were reared than was true of other physicians. Because of interruptions in their college education, they also tended to be older at college graduation. Their medical education more frequently occurred in less selective public medical schools with relatively low fees and those that were located in the Southeast, Plains or Western U.S.

Physicians in non-metropolitan areas, most of whom are in general/family practice, were readily distinguishable among the location groups. They were found to be very different from physicians in core counties of greater SMSAs and were especially different from those in fringe counties (suburban areas). The physicians practicing in counties having less than 20,000 residents in urban places were more frequently from small community backgrounds, as expected. However, many other characteristics were found to make an additional, significant contribution to distinguishing among these location groups. The rural practitioners had a higher interest in skilled trades, labor, farming, hunting and fishing activities in high school, a higher level of information on the Bible and on hunting and fishing, and a lower level of information on the arts. They were less definite about their future occupational choice in high school, had lower mathematics achievement and were more likely than future urban practitioners to have been in a non-college preparatory, less academic curriculum. They were also more likely to be Protestant, less likely to be Jewish, more likely to have fathers who were unskilled workers and not managers or businessmen, but just as likely to have fathers who were upper-level professionals (physicians, clergy, etc.).

These rural practitioners were less likely to have attended high school, college or medical school in New England or the Mideast than physicians who subsequently located in larger cities, and were relatively more likely to have lived in the Plains or Southeast. They did not go to highly selective colleges or colleges in large communities, they were more likely to have begun college relatively late, and their MCAT scores were lower than the average, with the greatest difference being in their verbal scores and the least in their quantitative scores. Moreover, they were much more likely to have attended less selective, public medical schools with relatively low tuition, fewer out-of-state students and a relatively small number of hospital beds.

Discriminant analyses showed that it was possible to discriminate among nine different specialty groups, three levels of specialization, and six community size groups, with a high level of confidence. Virtually every group in each outcome dimension could be significantly discriminated from every other group in that outcome dimension ($p \leq .01$) using only the TALENT data available in high school. The discriminations were improved by a moderate amount when college and medical school characteristics were added into the analyses. Further discriminant analyses were performed on a reduced set of six specialty groups formed by combining the general internists and pediatricians (the equivalent of general medicine other than GP/FP), the anesthesiologists, pathologists, and radiologists (support specialties), the psychiatrists and medical specialists, and the obstetricians and general surgeons, and by separating out the surgical specialists and subspecialists. The resultant discriminant functions correctly predicted the individuals' specialty group for one third (33%) of the cases, overall, using only high-school-age characteristics and for 38%, overall, using college and medical school characteristics. An accuracy of about 17% would be expected on a random basis. Prediction was best for the GP/FPs--fifty-five percent (55%) were classified correctly on the basis of their high school and college characteristics alone, and nearly 60% were correctly predicted when medical school characteristics were included. Prediction was least accurate in the case of the general internists and pediatricians, the numerically largest specialty group. Only 27% were correctly classified, as compared with 36% of the Ob-Gyns and general surgeons, 40% of those in medical specialties, 44% of those in surgical specialties, and 39% of those in support specialties.

Practice location was predicted correctly for 32% of the cases overall using only high-school-level data, for 36% of the cases when college data were included, and for 38% of the cases when data through medical school were included. A chance expectation would be 17% correctly predicted. Both extremes, large metropolitan fringe (suburban) and non-metropolitan practitioners, were classified correctly 44% of the time using only data through high school and 52% of the time when data through medical school were included.

A subsequent study has now been funded by the National Center for Health Services Research to integrate the findings of the TALENT study with those of the AAMC Longitudinal study, both of which were supported by NCHSR, and with other previous research.

Medical School Selection -- A Twenty Year Process

Sandra R. Wilson, Ph.D., Principal Research Scientist and Director, Health Research Group, American Institutes for Research, Palo Alto, CA

Given that the unique characteristics of future family practitioners and rural physicians were identifiable at points prior to medical school, a second objective of the study was to estimate (1) the proportion of individuals of similar characteristics who were present in the medical school applicant population for this cohort, and (2) the proportion of similar individuals who were present in the high school population. This was done using the discriminant functions (indices of potential) derived on the physicians. To make such estimates meaningful, the applicant and high school populations had to be restricted to those who had the abilities and interests that would make it at all reasonable to consider them to be potential physicians. For the applicants this was accomplished by examining the medical school "acceptance ratio" (ratio of successful/unsuccessful applicants) as a function of applicant MCAT test scores and choosing cutoff scores on the Verbal, Quantitative and Science tests at points (scores of 380, 420, and 420, respectively) where the acceptance ratio was increasing rapidly and where only 8% of the actual physicians but 37% of the unsuccessful applicants were excluded. For the sample of high school students, the TALENT General Academic Aptitude (GAA) composite test score was used. A cutoff was selected that was low enough to obtain relatively conservative estimates, but not so high as to exclude many of those who went on to enter medical school. The resultant cutoff included only the top quartile of the entire high school population and excluded only 5% of the subsequent physicians. Future physicians are not only distinguished from their high school classmates by their high level of academic ability but by a relatively high level of interest in the biological and the physical sciences. Therefore, the distribution of high school science interest scores for the physicians was examined and cutoff scores were defined such that all but 3% of the actual physicians scored above the cutoff. This cutoff was then applied to the high school students in the top quartile on GAA to further restrict the sample to those having a reasonable level of interest in the sciences. Checks were applied at each point to insure that the cutoff scores were appropriate in the sense that they did not exclude a disproportionate number of the actual GP/FPs or of the female or the black physicians. The resultant "potential physician" pool contained only 16% of the total high school student population (two thirds of those in the top academic quartile). The appropriate discriminant functions were then applied to the restricted applicant and the restricted high school samples to generate predicted specialties and practice locations for each of these "potential physicians." Attention was focused on the proportions of predicted GP/FPs and rural physicians in the high school and applicant populations and on their educational and occupational patterns after high school.

It was found that, whereas about 8% of the actual physicians in the cohort entered general or family practice, approximately 16% of the physicians in this cohort were classified as most like the GP/FPs on the basis of their characteristics through college. That is, they were more similar to the average of the characteristics of the actual GP/FPs than to those of any other specialty group. Among the high school students, 31.5% were classified as potential GP/FPs. The implication of these results is not that there could or should have been that many general/family practitioners trained. It is that among high ability high school students with an interest in medicine and other science careers, there was more than twice as high a proportion who resembled future GP/FPs than there was among those individuals who ultimately applied to medical school. The applicant pool, on the other hand, contained about the same proportion of potential GP/FPs as did the entrant pool which, of course, still means a numerically larger number of potential GP/FPs than actually entered medical school. Twice as many potential GP/FPs entered medical school as ended up in that specialty area, but those who were lost appear to have most often entered other primary care specialties--general internal medicine or pediatrics. (For this cohort, residencies in Family Practice were not as widely available as at present, so this loss is not surprising.) Thus while there appears to have been a loss of potential GP/FPs in medical school and/or graduate medical training, the loss was much greater between high school and the point of application to medical school. No major difference in the proportions of potential GP/FPs was seen between the academically qualified applicant and entrant populations, indicating that there was no systematic screening out of potential GP/FPs in the medical school selection process.

Rural physicians (those in counties of <20,000 urban population) comprise about 6% of physicians in this cohort who are now in practice. About 15% of the medical school entrants were classified as potential rural physicians on the basis of their high school and college characteristics. Among high ability high school students with science interests, 70% resembled rural physicians--31% most closely resembled those physicians who located in the smallest non-urbanized non-metropolitan communities and 39% those in urbanized non-metropolitan areas.

Where, then, were potential GP/FPs and the substantial number of potential rural physicians "lost" to the applicant pool, and what paths did they pursue after high school? To understand this it is necessary to first consider the fate of all of the "potential physicians" in the high school cohort. Approximately 88% of these students entered college, but 12% did not do so. Among those who entered college, only a very small percentage (3.2%) applied to medical school. The rest presumably either had a stronger interest in some other field or somehow dropped out of the competition. Between entrance into college and application to medical school, the student's selection of a major field is a key decision and one that is strongly related to the probability of applying to medical school for students anticipating a medical career; this major field is presumably in part a reflection of personal interests and in part a result of what the student perceives to be necessary to gain admission to medical school. Of the 1.2% who chose a "premed" major, about half applied to medical school; of the 6% who chose a biological science major, 20% applied; of 6% in the physical sciences, only 10% applied; of those in math/engineering, in social science majors, other health professions majors, and other or unknown majors, even smaller proportions applied.

Compared with the potential physicians as a whole, the potential GP/FPs were much less likely to make it into the applicant pool. Nearly one fifth did not enter college. Of those who did enter college, about the same proportion entered premed or science majors as among all the potential physicians, but only about half as many of the GP/FP science majors applied to medical school. As a result, only 1.6% of the potential GP/FPs in the high school sample entered the applicant pool, as compared with 3.4% of the non-GP/FPs (2.8% overall for potential physicians). A similar picture was found for the potential rural physicians, most of whom are GP/FPs. The proportion of potential rural physicians who did not attend college was 15%. Proportionately fewer chose a premed major, although the proportions choosing a science major were similar to the proportions of the total potential physician group. Only 1.4% entered the applicant pool.

Among women and blacks, the loss of potential physicians occurred in a somewhat different manner. Among those classified as potential physicians on the basis of academic abilities and interests, 42% were women. Nearly 20% of these women did not attend college, compared with about 6% of the men. Fewer of the women who entered college pursued science majors, and of those who did (including female premeds) fewer applied to medical school. Overall, only 0.7% of the potential female physicians entered the applicant pool, as compared with 2.8% of all potential physicians.

Among potential black physicians, 94% entered college as compared with 88% entering college among all potential physicians. (The average academic aptitude of the potential black physicians defined by the procedures used here is higher than the average aptitude of the majority group due to the differing slopes of the distribution of aptitude scores above the cutoff point for the two groups and this may explain the differential rate of college enrollment). The primary loss of potential black physicians after high school was related to the heavier choice of non-science majors by blacks. Of the small number who did choose science majors, the proportions who entered the applicant pool were as high or higher than the corresponding percentages for potential physicians as a whole; of those who chose non-science majors, the proportions of medical school applicants were below the overall average. The net effect of all of these factors was that only 1.8% of the potential black physicians entered the applicant pool. Overall, however, the most substantial loss of potential black physicians occurred before these students even reached high school.

The implications of these findings for attempts to redistribute medical manpower and for recruitment and selection of medical students will be discussed in this final presentation in the symposium.

SYMPOSIUM

MEDICAL SCHOOL LEARNING ENVIRONMENT: PROSPECTS FOR NEW RESEARCH DIRECTIONS

Chairman: Arthur Rothman, Ph.D.

Organizer: Sarah M. Dinham, Ph.D.

Participants: Sarah M. Dinham, Ph.D.
Christel A. Woodward, Ph.D.
T. Joseph Sheehan, Ph.D.

Critic: Edwin B. Hutchins, Ph.D.

MEDICAL SCHOOL LEARNING ENVIRONMENT: PROSPECTS FOR NEW RESEARCH DIRECTIONS

Sarah M. Dinham, Organizer, University of Arizona

Students know it...
Physicians know it...
Even curriculum evaluators know it.

Concerned educators have long known the importance of learning environment in higher education (Anderson and Walberg, 1974; Bloom, 1964; Pace and Stern, 1958; Plough and Dressel, 1973; Sheehan, 1969). Medical social scientists observe, colleagues discuss, faculties change, and researchers compare learning environments, but serious attempts to capture the subtleties of medical learning settings and to study them closely have been varied indeed.

This symposium's overall objective is to draw together the diverse views, research, and methodological issues relevant to the study of the medical school learning environment. To meet this objective, the symposium will:

- -review briefly the history of medical school learning environment measurement and research
- -present several viewpoints on the importance of the learning environment in medical education
- -present several views of the problems in studying medical school learning environment within and among medical schools
- -review several past and current research findings
- -present several viewpoints on the prospects for valid and useful measurement of learning environment variables and for future research on learning environment

The participants in this symposium agree that in the undergraduate medical curriculum, the learning environment (in its broadest conceptualization) may well be the single most important variable both within and among schools. Beyond agreement on this principle, however, the participants vary in:

- -their views of what constitutes "medical school learning environment"
- -their theoretical orientation toward research on this topic
- -their views of the first medical school environment measurement instrument and its descendents
- -the types of research they have conducted on the medical school learning environment
- -their optimism about the future for significant research concerning this topic.

This variance among the participants forms the backdrop for the moderator's introductory remarks, the three presentations, and the commentary of a critic whose work with this variable began over 20 years ago. The symposium will emphasize these disparate viewpoints, and will foster discussion of the prospects for capturing and studying this elusive and yet critical educational variable.

References

Anderson, G.J. and Walberg, H.J. Learning Environments. Chapter in H.J. Walberg (Ed.) Evaluating Educational Performance: A Sourcebook of Methods, Instruments and Examples. Berkeley: McCutcheon Publishing Co., 1974.

Bloom, B.S. Stability and Change in Human Characteristics. New York: Wiley, 1964.

Pace, C.R. and Stern, E.G., An Approach to the Measurement of College Environments. J. Educ. Psych. 40:269-277, 1958.

Plough, T. and Dressel, P., Assessing Academic Press. Improving College and University Teaching, 93-97, 1973.

Sheehan, T.J., Measuring School Environment: A New Approach. Proceedings: The 77th Annual Meeting of the American Psychological Association, Washington, A.P.A., 1969.

368

MEDICAL SCHOOL ENVIRONMENT MEASUREMENT FOR CURRICULUM ASSESSMENT

Sarah M. Dinham, Ph.D.
Office of Medical Education, The University of Arizona

Particularly in recent years, many and rapid changes in medical curricula have produced varied attempts to document curriculum change effects. It would have seemed that as important a factor as the learning environment would figure prominently in these evaluations. However, such criteria as scores on national certifying examinations and graduates' specialty choices are more often used; only infrequently are more subtle variables such as learning environment assessed (Beran, 1976; Dinham, 1976; Garrard and Weber, 1974).

This presentation reviews the development of medical school learning environment measurement, describes one school's findings and their methodological implications, and offers comments on the future for research and evaluation on medical learning environment.

The earliest of the successful medical school measures was the Medical School Environment Inventory (MSEI), used in the Association of American Medical Colleges' Longitudinal Study of the Class of 1960 to describe differences among schools (Hutchins, 1961). In the years after the MSEI's development, no fewer than five measures, and probably many more, were developed for assessing the learning environment of the medical school (Rothman and Ayoade, 1970; Hamberg and Dohner, 1970; Marshall, 1978; Johnson, 1978, Levy, et al., 1973). Most of these were developed for within-school use because the MSEI was intended for discerning differences among schools. As an alternative strategy, however, it might have been possible to investigate within-school applications of the original MSEI, especially for disparate programs or groups within the same school. This approach was used in curriculum studies at the University of Arizona.

Arizona's four medical learning environment studies were conducted as part of a larger evaluation addressing a major curriculum change. The first Arizona study was similar to Rothman's study and another conducted at the Chicago Medical School (Alberti, 1980). Entering medical students reported their expectations for the medical school learning environment, and then after the basic sciences ended were again administered the MSEI. Hotelling T^2 analysis was conducted to test for a composite difference and for differences on the six colinear MSEI factor scores. The T^2 was 79.59 ($F=12.25$; $df=6, 60$; $p=.0001$). New freshman medical students expected their school environment to show greater focus on academics, higher student motivation (both from faculty and within students), broader interests, and more specific (organized, detailed) instruction. They also expected a more generally positive atmosphere than they later reported they found. These results were similar to Rothman's, where students' expectations at matriculation differed on all scales from their mid-freshman year perceptions.

The second Arizona study compared the last class in a traditional four-year curriculum with the first class graduating from the newly designed three-year program. These two classes had different basic science programs but their clerkship and elective experiences coincided. The MSEI was administered to each class at the end of its basic science program; Hotelling's T^2 was 40.36 ($F=6.45$; $df=6, 117$; $p \leq .0001$). On four of the MSEI factors the groups differed ($p \leq .05$), with the students in the new three-year program reporting their basic science learning environment to be in general better, and in particular to evidence more academic interest, more faculty-induced motivation, and more organized, detailed instruction than the learning environment reported by the four-year traditional-program students.

The third Arizona study concerned only students enrolled in the new flexible-time three-year program. The study compared students making normal (three-year) progress toward their M.D. degree with students matriculating in the same class but subsequently extending their basic science programs from 15 to 24 months. The MSEI was administered midway through the basic sciences, when all students were feeling the intensity of the basic science program; some were coping adequately but others were deciding to lessen their course loads in favor of a four-year program. The Hotelling T^2 was 18.47 ($F=2.85$; $df=6, 62$; $p \leq .02$). Two factors differed between these groups at .10; the students extending their programs viewed their basic science environment as more faculty-motivated and as showing greater breadth of interest (e.g. in social as well as biological sciences) than did the students persisting in the three-year compressed basic science program.

The fourth Arizona study compared students' reports at the end of the basic sciences with their reports at the end of their M.D. program. The T^2 was 11.72 ($F=1.71$; $df=6, 35$; $p \leq .15$). This finding surprises those who assert that enduring the basic sciences is worthwhile because the clinical experiences will provide an improved learning environment.

In addition to the fact that the Arizona findings confirmed three of the original four hypotheses, the results case some interesting light on use of the Medical School Environment Inventory in monitoring curriculum change. Beyond its original intent for discriminating between schools, the MSEI promises some hope for use in distinguishing among curricular programs within a school. Despite the instrument's intent to maximize between-school and minimize within-school variance, the MSEI's sensitivity to curricular differences at Arizona suggests that a changed curriculum is not merely "same actors, different script" but might truly be a different educational experience. Further, the possibilities for research on inter- and intra-school differences are multiplied by the finding that this instrument has within-school as well as between-school measurement validity. Possibilities for concurrent validation with Rothman's Learning Environment Questionnaire are also intriguing.

But will further studies ever be done? The spattering of past research shows idiosyncratic and locally opportune research rather than the sequenced, coordinated building blocks of a logically developing line of inquiry. Medical school environments may be obvious, and schools may differ like night and day, but has medical school environment assessment been a passing fancy? It is a dream of the methodologists who believe that a variable as important as school environment must necessarily be measurable?

The Arizona studies suggest some hope for the study of medical school learning environment. So too do results from McMaster and Connecticut and elsewhere.

The challenges may well be both methodological and political, however, and the answers may come from arenas far beyond our own.

References

Alberti, J. Personal communication, 1980.

Barbee, R.A. and Dinham, S.M. Student Decision Making and Performance in a Flexible Time Curriculum. J. Med. Educ., 52:882-887, 1977.

Beran, R., The AAMC Study of Three-Year Undergraduate Medical School Curricula. Proceedings: The Fifteenth Annual Conference on Research in Medical Education, Washington: AAMC, 1976.

Dinham, S.M., Comparisons of Three- and Four-Year Students: Student Characteristics and Undergraduate Performance. Proceedings: The Fifteenth Annual Conference on Research in Medical Education, Washington: AAMC, 1976.

Garrard, J. and Weber, R.G., Comparison of Three- and Four-Year Medical Graduates. J. Med. Educ., 49:547-553, 1974.

Hamberg, R.L. and Dohner, C.W., Comparison of Student and Faculty Perceptions of the Medical School Environment. Proceedings: The Ninth Annual Conference on Research in Medical Education, Washington: AAMC, 1970.

Hutchins, E.B. The 1960 Medical School Graduate: His Perception of His Faculty, Peers, and Environment. J. Med. Educ., 36:322-329, 1961.

Levy, M., Morse, P.K., Leikelt, R.A., Dallman, J.J., and McDonald, T.F., Use of the Learning Environment Questionnaire to Assess Curricular Change. J. Med. Educ., 48:840-845, 1973.

Marshall, R.E., Measuring the Medical School Learning Environment, J. Med. Educ., 52:98-103, 1978.

Rothman, A.I. and Ayoade, F., The Development of a Learning Environment Questionnaire for Use in Curriculum Evaluation. J. Med. Educ., 45:754-759 (1970).

MEASURING LEARNING CLIMATE AT McMASTER: CONCEPTUALIZATION, OUTCOME AND CHALLENGE

Christel A. Woodward, Ph.D.
Programme for Educational Development, McMaster University

The term "medical school learning environment" may be as elusive as the often-debated term, "quality of care." There is little consensus on appropriate definitions for the educational environment of a medical school, and much less on appropriate research strategies for its study.

As with an institution's quality of care, a school's learning environment may be examined by scrutiny of its structure, its learning processes, or its outcomes. An examination of structure could include description of such elements as the components of the curriculum (both in terms of time allocation and types of activity), the student/teacher ratio, the actual curricular content, qualifications of faculty, and the philosophical basis of the school's educational approach. The examination of learning processes can include evaluation of the educational philosophy's implementation, and attitudinal questions such as how the participants perceive the educational program and how they value the educational experience. And finally, the learning environment's outcomes can be assessed through such variables as graduates' satisfaction, performance on national examinations, and career choices, as well as faculty satisfaction and factors in faculty attrition.

McMaster University Medical School's six year assessment of the learning environment focused on the second of these three approaches to learning environment. It was clear that the "environment" could not be captured adequately by mere description of its structural elements, and the outcomes of the medical program, while important in their own right, were not sufficient indicators of the learning environment. The McMaster study emphasized the learning processes observable in its medical program.

Since 1975, final year medical students at McMaster have described their perceptions of the medical school learning environment by reporting the extent of their agreement with a series of statements concerning aspects of the learning environment. The initial scale consisted of 36 items and was modeled on established instruments (Hutchins, 1961; Rothman and Ayoade, 1970). It grew and contracted from year to year, varying from 36 to 64 items as questions were deleted and others added. Response rate to the survey varied from 72% to 90% of graduating seniors across the years. In 1978 and 1979, approximately 80% of final year students completed the 50-item questionnaire.

The routine collection of student perceptions of their learning environment has been abandoned at McMaster after six years of study. This paper follows an initial review of learning climate evaluation findings by concentrating on the reasons for this decision in the light of the six years' study's outcomes, and raises questions about whether differences in students' perceptions of their environment within and across schools can be systematically studies in a sustained way.

Why was measurement of learning climate abandoned at McMaster? First, it was found that the data were only marginally useful. In particular, data from questions answered almost unanimously proved to have little information value. Given the limited number of questions that can be asked in a senior survey, uniform responses to "Students here are highly motivated" are not especially informative, even though the responses may be gratifying. Priority should be placed on questions about recent or contemplated program changes in order for the information to be useful to program planners. Second, it was found that data concerning one type of medical program, McMaster's, had little comparability with information from other programs.

If studying the climate for learning across schools with divergent educational philosophies is important, a major re-examination of the types of questions to be posed is needed. The challenge remains to develop an instrument to allow meaningful comparison across medical schools differing substantially in curricular organization and educational philosophy. Hutchins' Medical School Environment Inventory, useful in an era of relative homogeneity in medical education, can no longer adequately capture the diversity of medical school curricula. Equally challenging would be the development of an item pool with high information value for both within-school needs and among-school comparisons. Without such an instrument, measurement of educational climate will remain a sporadic activity adding little to our overall understanding of how the process of medical education influences its outcomes.

References

Hutchins, E.B. The 1960 Medical School Graduate: His Perception of His Faculty, Peers, and Environment. J. Med. Educ., 36:322-329, 1961.

Rothman, A.I. and Ayoade, F., The Development of a Learning Environment Questionnaire for Use in Curriculum Evaluation. J. Med. Educ., 45:754-759, 1970.

EGO DEVELOPMENT, FACULTY DEVELOPMENT AND EDUCATIONAL ENVIRONMENT

T. Joseph Sheehan, Ph.D.
Research in Health Education, University of Connecticut Health Center

I hope to show how a well-supported theory of personal development can be used to broaden and deepen our conception of faculty development and then to suggest that measurements of educational environments be examined for evidence of growth or decay.

First, what is the relationship between ego development and faculty development? In an earlier description (Sheehan, 1978, 1979) of Loevinger's stage theory of ego development (1976), I suggested that her stages could be used to describe the developmental level of a faculty. Faculty development programs are designed to improve teaching skills: Their deeper purpose is to improve the quality of instruction and to elevate the value of instructional activities in the minds of our faculty.

Loevinger's theory of ego development is a general framework for considering how we are developing as people. According to Loevinger, people develop through a sequence of stages. Those at the lowest stage recognize no rules and act strictly out of impulse. At the self-protective stage, which comes next, individuals do recognize rules, but the most important rule is "Don't get caught!" Interpersonal relationships at the self-protective stage are manipulative and exploitive. There is a conscious preoccupation with control and domination.

The conformist stage is next. Rules are recognized and obeyed, not so much from fear of punishment, but in hope of peer approval or in deference to peer expectations. Conformity is not to be confused with the appearance of being conventional.

Those at the higher stages, the conscientious, autonomous or integrated stages, have values that come from within themselves. Their actions are based upon inner convictions, and at times these actions may run counter to peer expectations or even counter to the law.

How might these stages describe the developmental level of a faculty? Where a faculty is primarily self-protective, it looks to its own interests and stakes. Territoriality is important. The reward system is important. "What's in it for me?" is the predominant attitude. At this level it is difficult for a faculty to do anything educationally meaningful. Even if one can bring about curriculum reform, it will not be lasting unless the faculty can rise above its self-protective instincts. Parenthetically, a faculty may operate at a self-protective stage because its leaders do also, and transmit a clear message through policy and administrative decisions that self-protection is fully justified and necessary for survival.

A forceful dean can probably persuade a conformist faculty to accept change, if only to keep up with the latest fads. However, a conformist faculty is likely to discard such changes as soon as the tune from the bandwagon changes.

A faculty must be at the conscientious stage or above before it can be truly effective. Some inner belief is driving the faculty, this internal force goes beyond self-interest and beyond what others expect. There is a sense of mission which draws faculty from the security of departmental nests toward a deeper commitment to the mission of the institution.

Besides providing this more profound view of faculty development, Loevinger's theory also enriches our views of educational leadership. The effective educational leaders are canny at reading the developmental level of their faculty. They lead faculty to higher levels in a number of ways: by reassuring the self-protective, by sharing responsibility with the conformist, and by inspiring everyone with a sense of mission and enthusiasm for that mission. They recognize the developmental level of the faculty and help them to grow by appealing to them with reasoning one stage beyond where they are. Step by step, they raise their faculty to the highest level they are capable of reaching. They are also conscious of all the pressures that can cause backsliding: budget restraints, space, teaching loads, senior appointments and promotion problems.

Since it is doubtful that all educational leaders can do what the good ones do intuitively, my thesis is that environmental measures can be used to monitor where the faculty is developmentally, and to monitor the kinds of perceptions they are creating in the minds of students. Although the findings are now dated, I would offer an earlier study (Sheehan, 1970) as a model of how an environmental measure can be used to detect comparative faculty-student views of the learning environment.

I believe that a faculty's developmental level, its attitudes toward education and toward student development are the major determinants of the educational environment. I also believe that the details of how the curriculum is organized are less important than how these attitudes and beliefs are perceived by students. As Stern pointed out long ago (1962), some environments are designed to stimulate, while others to satisfy. Which is better? Pearls come from aggravated oysters, but you can only get milk from contented cows.

References

Loevinger, J. Ego Development: Conceptions and Theories, San Francisco, Jossey-Bass, 1976.

Sheehan, T.J. "Medical School Climate: Comparative Faculty and Student Views," Journal of Medical Education, 45:880-887, 1970.

Sheehan, T.J. "Economics of Curriculum Change," Proceedings of the Seventh Symposium on Veterinary Medical Education, Purdue University, pp. 58-63, 1978.

Sheehan, T.J. "Faculty Development--Another Meaning," Journal of Medical Education, 54:255, 1979.

Stern, G. "Environments for Learning," in The American College: A Psychological and Social Interpretation of the Higher Learning (N. Sanford, ed.), New York: John Wiley and Sons, 1962.

SYMPOSIUM

LEARNING DURING THE SENIOR CLERKSHIP

Chairman: Hugh M. Scott, M.D.

Participants: R. Kennedy Smiley, M.D.
Hugh M. Scott, M.D.
Vimla Patel

Discussants: W. Dale Dauphinee, M.D.

376

Learning During the Senior Clerkship

Objective

The Symposium will examine both the methodological issues involved in studies of student learning during Senior Clerkships as well as the results of three completed representative studies. Discussion will then focus on possible explanations for the differences found in these studies, and the implications of results obtained. Finally, directions for future research will be identified.

Introduction

The Senior Clerkship (or sub-internship as it is sometimes called) is the culmination of the undergraduate medical experience. It is an educational format somewhat unique to Medicine. As such it is somewhat surprising that relatively little attention has been paid to it in the Medical Education literature.

This certainly does not reflect any lack of questions or controversy surrounding the optimum methods of conducting this vital educational experience. Instead, opinions abound concerning such issues as the optimum number and length of rotations; their optimum sequence; the role of electives. Indeed, some have even questioned what learning does occur, is there learning in all areas of faculties objectives, or just in one, say psychomotor skills. Finally, there is even concern that this experience may be to the detriment of some objectives in, for example, the affective domain.

This Symposium will present three studies which have addressed some of these questions and then consider their implications in the broader context of future research directions.

The first study examines the long debated question of sequence, using Internal Medicine as the "target discipline". The second touches as well on sequence (with perhaps partially conflicting results), and continues further to examine the question of factual learning versus problem-solving or "clinical" learning, again using Internal Medicine as the reference discipline. The third study expands the examination to learning in domains other than cognitive, and includes an evaluation of the learning environment in a clerkship setting. As well, the disciplines of Surgery and Pediatrics are included to see if any findings are "discipline specific".

Student Performance in Internal Medicine: The Effect of Previous clerkship experience

R. Kennedy Smiley
Professor and Head
Department of Medicine
University of Ottawa

To test the assumption that students who have completed other Clerkships are likely to perform better in Internal Medicine, a prospective study of 160 senior students from two classes was undertaken. Groups of students with little or no Senior Clerkship experience before their Internal Medicine Clerkship were compared with groups of students who had extensive Clerkship experience in disciplines other than Medicine. Groups were compared with respect to ward-performance scores and scores on identical multiple-choice examinations. Neither ward-performance nor multiple-choice scores revealed differences attributable to previous clerkship experience. It is concluded that despite areas of instruction and experience which overlap between Medicine and other disciplines, previous Clerkship experience apparently does not have a beneficial effect on student performance in Internal Medicine, as measured by these two commonly used methods of evaluation.

378

The Learning of Facts and Clinical Problem solving skills during the Senior Clerkship in Medicine

Hugh M. Scott, Professor and Chairman,
Department of Medicine
University of Sherbrooke

Before McGill University moved to change its curriculum to the former partially described in the format partially described in the next paper, a study was undertaken to estimate learning in Medicine during the previous twelve month senior clerkship. It attempted to answer the following questions -

i) Were there significant learning gains in both factual knowledge (as measured by National Board multiple choice examinations) and Clinical Problem solving skills (as measured by National Board Patient Management Problems) during the three month Medicine Clerkship?

ii) Did the level of ability in both categories increase at the time of entry as the year progressed?

iii) If differences were demonstrated did they persist to the end of the year?

The results demonstrated that there were significant learning gains in both categories in all four clerkship periods. As well, entry or "pre-test" scores were higher for those groups who had had their Surgery, Pediatrics and Obstetrics Clerkships before Medicine. However, a preceding elective did not enhance, and may have even decreased entry scores.

Finally, and perhaps fortunately, all groups had equal performance in a further examination of both categories taken at the very end of the year as part of a certifying examination.

A Multi-faceted Approach to Student Learning During the Senior Clerkship

Vimla Patel
Research Associate
Centre for Medical Education
McGill University
Montreal, Quebec

As part of a continuing comprehensive evaluation of the revised undergraduate curriculum at McGill, particular attention has been paid to the final clinical or "clerkship" phase. It was concluded that no single measure could properly describe the degree of attainment of the multiple important objectives outlined by the curriculum's designers for this stage in the program. Therefore, this paper will describe the overall evaluation strategy as well as specific tools and techniques used to evaluate i) learning in cognitive, psychomotor and affective domains during three major disciplinary clerkships (medicine, surgery and pediatrics) ii) what, if any, role does discipline sequence play in the above iii) what are the parameters of time utilization by students in the program. The overall structure of the evaluation design is given in Figure 1 below.

The results show that the learning of 1) factual knowledge as part of the cognitive domain is greatest in medicine, followed by pediatrics and least in surgery, 2) interpersonal skills is greatest in pediatrics, followed by medicine, and least in surgery.

Furthermore, the factual knowledge scores in 1) medicine are shown to be enhanced by previous clerkship experience in pediatrics and surgery; 2) pediatrics are shown to be enhanced by previous clerkship experience in medicine and surgery; 3) surgery are not affected by the previous experience in medicine and pediatrics. The learning environments in different rotations are found to be fairly constant. The results further indicate that there is a transfer of interpersonal skill from one discipline to another.

Learning in the affective domain did occur during the clerkship year. However, it is a source of reflection that even after this learning gain, scores did not achieve the level of a first year student control group. In other words, it appears that a loss has occurred between the end of first year and the beginning of the clerkship that could not be totally made up.

Fig. 1 Instrumentation in Evaluation Design

DOMAINS		INSTRUMENTS
(1) COGNITIVE	Factual Knowledge	Multiple Choice Questions
	Problem Solving	Patients Chart Review
(2) PSYCHOMOTOR	Clinical Skill	Physical Examination Assessed Using a) Clinical Checklist b) Questionnaire by Patients
	Inter-personal Skill	Videotapes of Doctor-Patient Interviews
(3) AFFECTIVE	Attitude towards health care	Questionnaire
(4) ORGANIZATION	Student time	Diary
	Learning Environment	Questionnaire

Learning During the Senior Clerkship

W.D. Dauphinee
Associate Dean (Medical Education) and
Director, Centre for Medical Education
Faculty of Medicine
McGill University
Montreal, Quebec

The quality and quantity of experiences during the senior clerkship, (clinical clerkship, sub-intern) have come under increasing surveillance recently. Medical schools are concerned about the effectiveness of the curricular changes that were introduced in the 1970's. Licensing bodies (in Canada particularly) have become concerned about the content of the senior clerkship. Despite these concerns, there is a paucity of reports of either assessment of programs or student learning in the literature.

The studies of Smiley, Scott and Patel have focused on the effect of the senior clerkship on student learning and student performance. Although the studies were done in schools only 120 miles apart, seemingly different results were obtained.

In reality, the results may not be contradictory. Patel classified student learning under four headings: cognitive, psychomotor, affective and organization. She looked at learning during three different clinical rotations: medicine, surgery and pediatrics at the start and end of the clerkship. She found learning varied form rotation to rotation for any given category of activities. Smiley utilized multiple-choice question and ward performance but found no differences that could be attributed to previous clerkship experience. Utilizing multiple choice examinations Scott found differences that could be attributed to previous clerkship experience. One might argue that Smiley may have found differences if he had looked at specific categories of performance similar to Patel's. An alternative explanation might be that the objectives and activities of individual rotations at Ottawa may differ from those at McGill - and this is reflected in the students' learning. An accounting of the activities of senior clerks might reflect such priorities.

The two studies raise some methodological issues: the need for pre-tests, the need to test the reliability of instruments and the possible short-comings of implicit overall ward evaluations as indicators of student learning or accomplishment.

What is for the future? The two studies give us several new directions for future studies of clinical clerkships. How do senior clerks spend their time and do their activities reflect the implicit or explicit goals of the rotations? Are some ward structures and organizational arrangements helpful or detrimental to clerkship performance? Do such organizational arrangements and teaching strategies work equally well for all clinical rotations? What factors determine the presence or absence of synergistic or additive learning effects between clinical desciplines? What effects do electives or other open-ended clinical clerkship experiences have on student learning and accomplishment? There is a need to look much more closely at each of the components of student learning. Do they function independently? Are some more appropriately taught in specific disciplines? What is the effect of medical subspeciality versus general medically oriented teaching units? Lastly, there is a great need to develop better instruments with which more reliable and valid measures of student learning can be accomplished.

Needless to say, the concept of the clerkship is not under fire. However, the time has come to use solid and reliable methods of assessing and evaluating the clerkship experience. New strategies in developing clerkship experiences should be based on the best possible evidence, and ideas developed from such evaluations and studies - not the swings of the pendulum of pedagogical fashion.

SYMPOSIUM

THE *STANDARDS* DEVELOPED BY THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION: REFLECTIONS ON THEIR IMPACT FOR EVALUATION IN MEDICAL EDUCATION

Chairman: Hulda Grobman, Ed.D.

Participants: Hulda Grobman, Ed.D.
Phillip Bashook, Ed.D.
Vic Neufeld, M.D.

Discussants: Joseph S. Gonnella, M.D.

384

SYMPOSIUM

THE STANDARDS DEVELOPED BY THE JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION: REFLECTIONS ON THEIR IMPACT FOR EVALUATION IN EDUCATION IN MEDICINE

Organizer: Hulda Grobman, Ed.D., School of Medicine,
St. Louis University

OBJECTIVES OF THE SYMPOSIUM:

The major objective is to explore the relevance and potential impact of the newly published *Standards for evaluations of educational programs, projects, and materials* (1) for medical education. The *Standards* were prepared by a 16-member Joint Committee on Standards for Educational Evaluation to complement the *Standards for educational and psychological tests* (2). If these new *Standards* are accepted by program developers, granting agencies, and consumers of evaluations in considering proposals and evaluation efforts in medical education, their impact will be profound.

The symposium will examine: (1) the orientation of the developers, consultants and pilot populations; (2) scope of the standards and intended use; and (3) appropriateness and relevance of the standards for evaluations in medical education, using as the context two case studies. The discussant will focus on application of the standards in these case studies and possible implications and consequences of using the standards in evaluation studies in medical education.

(1) Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects, and materials.* New York: McGraw Hill Book Company, 1980.

(2) Joint Committee of the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. *Standards for educational and psychological tests.* Washington, D.C.: American Psychological Association, 1974.

Hulda Grobman, Ed.D.
St. Louis University School of Medicine
St. Louis, Missouri

THE STANDARDS: THEIR GENESIS AND APPROACH TO PROGRAM, PROJECT AND MATERIALS EVALUATION

Educational evaluation as a recognized systematic field of endeavor is less than 20 years old. Particularly in the 1960's, but lasting into the 1970's, criteria used in evaluations and meta-evaluations reflected the scientific research model; test use reflected the criteria of the Standards for educational and psychological tests (1). As educators gained evaluation experience, credibility of evaluation as a legitimate endeavor, validity of non-traditional approaches, and need for basic evaluation theory gained recognition. The Standards, 1974 revision, identified a need for separate but complementary guidelines on program evaluations and test use for such evaluations.

A 16-member group appointed to work on this new task includes representatives of the American Psychological Association, American Educational Research Association and National Council on Measurement in Education, and of organizations in public school administration, teaching, governance, curriculum, and guidance and private education. Writers, project staff, reviewers, and field testers reflect a similar thrust. This Committee's Standards for evaluations (2), issued in summer 1980, provides philosophical and pragmatic frameworks for program, project and materials evaluations, presented as a developmental effort to be used as a general guide. The 30 standards are organized in sequential categories of: A. Utility; B. Feasibility; C. Propriety; and D. Accuracy. Each standard is examined in the context of: Overview, Guidelines, Pitfalls, Caveats (warnings against overzealous implementation) and Illustrative Case, to provide a concise delineation of that standard and its appropriate applications. The thrust may be best summarized by the checklist (Chart 1) provided in the Standards to assist evaluators in communicating how the standards were used, what judgments about meeting the standards were reached, and the evidence supporting such judgments.

Some insight into the standards may be gained by considering the personnel involved in development, and some of the problems they have faced in the role of evaluator. This may be reflected in the choice of illustrative cases, generally involving public education, and often using classic examples of fiascos in evaluation. There is major concern with political implications and for the "right to know" of various evaluation participants and diverse audiences. Evaluation is presented as a clearly defined, formalized effort - discrete from other operational activities, even when overlap of personnel exists - with the evaluator(s) clearly identified and the relationship of evaluator(s) and client formally delineated. Assignment of priorities to values is a deliberate and critical step in planning and implementation. Value conflicts are to be resolved non-passionately, based on priorities within and among categories; and the ability to provide

Citation Form*

CHART 1

The *Standards for Evaluations of Educational Programs, Projects, and Materials* guided the development of this (check one):

____ request for evaluation plan/design/proposal
____ evaluation plan/design/proposal
____ evaluation contract
____ evaluation report
____ other

To interpret the information provided on this form, the reader needs to refer to the full text of the standards as they appear in Joint Committee on Standards for Educational Evaluation, *Standards for Evaluations of Educational Programs, Projects, and Materials.* New York: McGraw-Hill, 1980.

The *Standards* were consulted and used as indicated in the table below (check as appropriate):

Descriptor	The Standard was deemed applicable and to the extent feasible was taken into account	The Standard was deemed applicable but could not be taken into account	The Standard was not deemed applicable	Exception was taken to the Standard
A1 Audience Identification				
A2 Evaluator Credibility				
A3 Information Scope and Selection				
A4 Valuational Interpretation				
A5 Report Clarity				
A6 Report Dissemination				
A7 Report Timeliness				
A8 Evaluation Impact				
B1 Practical Procedures				
B2 Political Viability				
B3 Cost Effectiveness				
C1 Formal Obligation				
C2 Conflict of Interest				
C3 Full and Frank Disclosure				
C4 Public's Right to Know				
C5 Rights of Human Subjects				
C6 Human Interactions				
C7 Balanced Reporting				
C8 Fiscal Responsibility				
D1 Object Identification				
D2 Context Analysis				
D3 Described Purposes and Procedures				
D4 Defensible Information Sources				
D5 Valid Measurement				
D6 Reliable Measurement				
D7 Systematic Data Control				
D8 Analysis of Quantitative Information				
D9 Analysis of Qualitative Information				
D10 Justified Conclusions				
D11 Objective Reporting				

Name ____ Date ____ Position or Title ____ Agency ____

Relation to Document (e.g., author of document, evaluation team leader, external auditor, internal auditor) ____

*The Publisher gives permission to copy this page.

an "objective" description of this priori[illegible] process and the need for such an overt description is implied.

Despite warnings that the Standards are formative, not all-encompassing, and not intended to discourage innovation, the format may encourage overacceptance with automatic, rule-of-thumb reference. This makes it imperative that persons involved with education in medicine examine the underlying philosophy of the Standards and the practical implications, before such acceptance occurs in their field. Attention should be given to such questions as: Are standards developed primarily in the context of public education appropriate/practical for medical education? Will they tend to stereotype evaluations? These are questions to be examined in the following papers.

(1) Joint Committee of the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. Standards for educational and psychological tests. Washington, D.C.: American Psychological Association, 1974.

(2) Joint Committee on Standards for Educational Evaluation. Standards for evaluations of educational programs, projects, and materials. New York: McGraw-Hill Book Company, 1980. Rev. Ed.

APPLICATION OF THE STANDARDS

TO THE EVALUATION OF THE FLEXNER REPORT: A CASE STUDY

Philip G. Bashook, Ed.D.
Educational Development Unit
Michael Reese Hospital and Medical Center

The Standards are claimed to be valuable in "evaluating an evaluation" (1, p.23). This presentation will address this issue by answering two questions.

1. How relevant are the Standards in reviewing an existing evaluation report, Abraham Flexner's 1910 Report (2).

2. Does applying the Standards help identify significant omissions or discrepancies in the Flexner Report?

The Flexner Report was formally published on June 5 or 6, 1910, as Bulletin Number 4 (1910) by the Carnegie Foundation. It was divided into two parts and included an appendix containing a summary of the findings for each of the 155 schools surveyed. A more complete description will be summarized and distributed as a handout.

Method

The Standards were considered using two criteria: relevance, and usefulness in identifying significant omissions or discrepancies. These observations serve as the basis of this presentation. In addition to the published Flexner Report(2) other data sources are reports from the same era, Flexner's autobiography(3), as well as medical historian's interpretation of events surrounding the Flexner Report. Attached is a selected bibliography containing examples. The complete bibliography will be provided as a handout.

Results

Described below are some standards considered helpful for evaluating the Flexner Report. Each standard will be discussed under the major headings used in the Standards document. The actual presentation will expand and include all relevant standards.

A. UTILITY

Standard A.2. Evaluator Credibility: Flexner was an educator not a physician. He was hired by Pritchett of the Carnegie Foundation to study the medical schools "not from the point of view of the practitioner but from the standpoint of the educator" (3, p.71). He seemed credible to the medical community because of Carnegie Foundation influence and secret support from the American Medical Association(4).

Standard A.3. Valuational Interpretation: Flexner omits from the report the underlying basis for his judgments about the medical schools. In his 1960 autobiography, 50 years after the report, he says he used the Johns Hopkins Medical School as his model for comparison (3, p.74).

Standard A.8. Evaluation Impact: There is no question that the Flexner Report has been credited with changing the course of medical education. Yet, the changes actually began years before the report was published(6).

B. FEASIBILITY

Standard B.1. Practical Procedures: Flexner visited each of the 155 medical schools either alone or with Colwell of the AMA. His visits lasted between 1-3 hours and resembled current site visit procedures used today for accrediting educational programs (3, p.74-75).

C. PROPRIETY

Standard C.7. Balanced Reporting: Flexner made scathing attacks at the more commercial medical schools and frequently used his visits as a weapon before making his report to change what he perceived as bad. When schools complied with his requests the reports were less harsh(2,3).

D. ACCURACY

Standard D.5. Valid Measurement: His visit to each school is claimed as the data source in the Report(2). However, Monahan(5) located questionnaires on school finances which Flexner sent to schools and must have used in preparing the report. Were there other unsubstantiated data sources?

Discussion

The Standards contain a "Functional Table of Contents" suggesting which standards should be most relevant for various evaluation functions. There are some serious problems with these suggestions. Also, the Standards seem to assume that a "good" evaluation must provide full and frank disclosure of data and judgments and quantitative data are always superior to qualitative data. Both assumptions will be challenged in this presentation.

---

1. Joint Committee on Standards for Educational Evaluation. Standards for Evaluation of Educational Programs, Projects, and Materials. New York: McGraw-Hill, 1980.

2. Flexner, Abraham. Medical Education in the United States and Canada. A report to the Carnegie Foundation for the advancement of teaching. Bulletin #4(1910), Boston: D.B. Updike, The Merrymount Press, 1910. (Reprint edition 1972, New York: Arno Press).

3. Flexner, Abraham. An Autobiography. New York: Simon and Schuster, 1960.

4. Bevan, Arthur. Cooperation in Medical Education and Medical Service--Functions of the Medical Profession, of the University of the Public, J.A.M.A. 90 (#15), 1173-1177, April 14, 1928.

5. Monahan, James. Medical Historian, Chicago, Illinois. Personal Communication, July 21 and July 25, 1978.

6. Flexner, Abraham. Medical Education--a Comparative Study. New York: McMillan Co., 1925.

APPLICATION OF THE STANDARDS TO THE EVALUATION OF THE McMASTER M.D. PROGRAM: A CASE STUDY

Vic Neufeld, M.D. and Christel Woodward, Ph.D.
Program for Educational Development
Faculty of Health Sciences, McMaster University

This presentation will address these questions:

(1) How relevant are the Standards in reviewing the efforts that are being made to evaluate the McMaster M.D. Program.

(2) Are there gaps in this activity which are identified as a result of applying the Standards.

The McMaster M.D. Program, which admitted its first students in 1969, had some unusual features in its admission process and in its curriculum. Within a few years, it became apparent that there was sufficient internal and external interest--to initiate a systematic evaluation of the McMaster approach. These evaluation activities have been described elsewhere, (1, 2) and will be summarized in a handout.

Method:

The Standards were read independently by two McMaster faculty members who are familiar with the M.D. Program evaluation activities. Each standard was considered, using two criteria: relevance, and usefulness in identifying gaps. These independent observations were then compared and combined for this presentation.

Results:

The following are some of the Standards which were thought to be particularly relevant to the evaluation of the M.D. Program. They are discussed briefly under each of the four headings used in the Standards document. These comments will be expanded in the actual presentation.

A. UTILITY:

A1 - Audience Identification: While there are both internal and external audiences, these have not been formally ranked.

A3 - Information Scope and Selection: Selection has been made on the basis of feasibility and interest. Are some important issues avoided as a result?

A4 - Valuational Interpretation: This standard is of particular interest in the use of national licensing and certifying examinations as part of the program evaluation procedure.

A8 - Evaluation Impact: The degree to which the internal McMaster "audience" has used the results to date will be reviewed.

B. FEASIBILITY:

B1 - Practical Procedures: The strategy of first, second and possibly third generation studies will be discussed.

B3 - Cost Effectiveness: Some planned studies have not been initiated because of cost. Have the smaller, fundable studies been worth the effort? Are they enough?

C. PROPRIETY:

C2 - Conflict of Interest: For a variety of reasons, virtually all the studies are conducted by "inside" evaluators. This raises questions of credibility and bias. The advantages and disadvantages of internal evaluation will be discussed.

C5 - Rights of Human Subjects: We will review our experience with "informed consent" procedures.

D. ACCURACY:

D1 - Object Identification: A major methodologic dilemma is the inability to clearly separate out the relative effects of the admissions process and the curriculum on graduate outcomes. Suitability for the curriculum is assessed as part of the admissions process.

D5 - Valid Measurement: Our experience with the development of new performance measures will be reviewed.

Discussion:

More than half of the Standards were relevant to our activities. They provided a focus for critical review of our evaluation efforts. However, it is unlikely that our overall evaluation plan will be substantially changed as a result of this review.

References:

1. Woodward, C.A., Neufeld, V.R.: "A Long Range Evaluation of Curriculum Innovation in Medical Education". In A.D. Hunt and L.E. Weeks (Eds.) Medical Education Since 1960: Marching to a Different Drummer. East Lansing, Michigan: The Michigan State University Foundation, 1979.

2. Neufeld, V., Woodward, C., Van Boven, C., Clarke, R., Segall, A.: "What Are The New Medical Schools Doing About Curriculum Evaluation: An International Perspective". A Symposium. Annu Conf Res Med Educ, 17, 1978, pp. 449-458.

APPENDIX

RIME CONFERENCE SCHEDULE

PRESENTATION OF PAPERS

STANDARDS AND STANDARD SETTING

MODERATOR: Bryce Templeton, M.D.
National Board of Medical Examiners

The Use of the Rasch Model to Test the Equivalence of Two Methods of Standard Setting
Peter H. Harasym, Ph.D.

A Comparison of Several Score Cutting Procedures and Their Effects on Success Rates
Ernest N. Skakun, Sam Kling

Content Representativeness and Student Performance on National Board Part I Special Subject Examinations
William C. McGaghie, Hugh J. Burford, Donna H. Harward

The Feasibility of Using a Canadian Examination for the Certification of Australasian Candidates in Psychiatry
Ernest N. Skakun, Donald R. Wilson

DATE: October 29, 1980
TIME: 9:00 a.m. - 10:45 a.m. LOCATION: Georgetown East

PREDICTING CAREER OUTCOMES

MODERATOR: Marcia Z. Wile, Ph.D.
Case Western Reserve University, School of Medicine

The Role of Personal Themes in Medical Specialty Choice
Tod S. Sloan, Sandra F. Bermann

Prediction of Medical Student Career Choice From a Freshman Personality Profile
Murray M. Kappelman, M.D., Bernice Sigman, M.D., Leslie Walker-Bartnick, M.A.

Proximate and Long-Term Effects of Early Exposure to Primary Care
Marian Osterweis, Ph.D., David L. Rabin, M.D., M.P.H.

The Impact of Residency on Physician Practice Patterns: An Exploratory Analysis of Young Internists
Peter A. Weil, Ph.D., Mary Kay Schleiter, M.A., Alvin R. Tarlov, M.D., Robert C. Mendenhall, M.S., Christy Moynihan, Ph.D.

Physician Career Satisfaction: Another Look
Betty Hosmer Mawardi, Ph.D.

DATE: October 29, 1980
TIME: 9:00 a.m. - 10:45 a.m. LOCATION: Georgetown West

PRESENTATION OF PAPERS

TEACHING CLINICAL SKILLS

MODERATOR: Edward Coppola, M.D.
Michigan State University, College of Medicine

Information Mapping In Introduction to Clinical Medicine
Emil Petrusa, Ph.D., Paula K. Horvatich, Ph.D., James C. Guckian, M.D.

Perceptions of Student-Patient Relations
Margaret E. Uguroglu, M.Ed., Rodney Nelson, M.D., Charles Kanakis, M.D.

A Comparison of Structured and Self-Directed Approaches to Teaching Interviewing and Interpersonal Skills to Pediatric Residents
Leslie S. Jewett, Ed.D.

Teaching Medical Interviewing Skills: A Comparison of Medical and Non-Medical Tutors
Julain Bird, M.R.C.P., M.R.C. Psych., Christopher D. Lorish, Ph.D., Stephen Cohen-Cole, M.S., C. Kirk Avent, M.D.

DATE: October 29, 1980
TIME: 9:00 a.m. - 10:45 a.m. LOCATION: Lincoln West

TEACHING IN THE CLINICAL SETTING

MODERATOR: John H. Littlefield, Ph.D.
University of Texas Health Sciences Center

Analysis of Clinical Experience
Collin Baker, M.D.

Ideal and Actual Resident Teaching Practices in a University Hospital
Carter Zeleznik, Ph.D., Paul Brucker, M.D.

Similarities of General Medicine Clinic in a Teaching Hospital to Internal Medicine Practice
Roberta A. Monson, M.D., Judith Jameson, M.P.H.

The Relationship Between Medical Student Clerkship Activities and Performance on NBME Part II
Judith G. Calhoun, Ph.D., Alan L. Hull, Ph.D., Wayne K. Davis, Ph.D.

DATE: October 29, 1980
TIME: 11:00 a.m. - 12:45 p.m. LOCATION: Georgetown East

PRESENTATION OF PAPERS

PLANNING CONTINUING MEDICAL EDUCATION

MODERATOR: Oscar A. Thorup, M.D.
University of Virginia, School of Medicine

The Educational Value of a Model Medical Care Evaluation Program
L.J. Sandlow, M.D., P.G. Bashook, Ed.D., J. Maxwell, M.A.

Characteristics Identified Upon Entrance to Medical School Associated With Future Participation in Professional Education
Linda K. Gunzberger, Ph.D.

Physicians Practice Profiles: A Comparison of Sampling Methods
Lynn Curry, Ph.D.

Using Medical Audit Results to Plan Continuing Medical Education in Community Hospitals
Charles E. Osborne, Ed.D.

DATE: October 29, 1980
TIME: 11:00 a.m. - 12:45 p.m. LOCATION: Lincoln West

IMPACT OF RESIDENCY PROGRAMS: MEASUREMENT PERSPECTIVES

MODERATOR: Kenneth W. Rowe, Jr., M.D.
University of Cincinnati, School of Medicine

Faculty Perceptions of American and Foreign Pediatric Residents
T. Joseph Sheehan, Ph.D., Susan D.R. Husted, Dan Candee, Ph.D., Charles D. Cook, M.D.

Performance on Part III of the National Boards - The Effect of Residency Training
Jon Veloski, M.S., Joseph S. Gonnella, M.D.

Patient Instructors as Evaluators of Housestaff Clinical Competence
Paul J. Rutala, M.D., Paula L. Stillman, M.D., Darrell L. Sabers, Ph.D.

DATE: October 29, 1980
TIME: 1:30 p.m. - 3:30 p.m. LOCATION: Georgetown East

PRESENTATION OF PAPERS

ENHANCING STUDENT/FACULTY ENVIRONMENT INTERACTIONS

MODERATOR: Marilyn Heins, M.D.
University of Arizona, College of Medicine

Medical Student Needs: What and When
Grant D. Miller, M.D., Elizabeth C. Miller, M.A., Owen C. Peck, M.D.

Affective Learning in Medical Education
Elsbeth Kahn, Ph.D., Sandra L. Lass, Ph.D., Russell Hartley, M.A., Helen Kornreich, M.D.

Evaluation of a Medical School Learning Environment
A. Harris, Ph.D.

The Effects of Group Study Skills Counseling and Applied Relaxation on Study Behaviors and Test Anxiety in Medical and Dental Students
David G. Schroeder, Ph.D.

Development, Implementation and Evaluation of a Program to Improve Lecture and Presentation Skills
Donald S. Masler, M.D., Patricia R. Austin, M.A., Ilene B. Harris, Ph.D.

DATE: October 29, 1980
TIME: 1:30 p.m. - 3:30 p.m. LOCATION: Georgetown West

EVALUATING CONTINUING MEDICAL EDUCATION

MODERATOR: Charles F. Johnson, M.D.
East Tennessee State University, College of Medicine

Investigations in CPR Training
David A. Gass, M.D., Lynn Curry, Ph.D.

Efficacy of Traditional Continuing Medical Education in Changing Physician Knowledge and Behavior in the Care of Patients with Acute Myocardial Infarction
Carl W. White, Janet L. Roseman, Mark A. Albanese, Donald D. Brown, Marcia K. Whitney, Richard M. Caplan

A Model Continuing Educational Delivery System for Isolated Physicians in the Area of Pulmonary Medicine: Development and Evaluation
Muriel H. Bagshaw, M.D., Wylie L. McNabb

Patient Care Appraisal in the Ambulatory Setting: Effectiveness as a Continuing Medical Education Tool
Wayne Putnam, M.D., Lynn Curry, Ph.D.

DATE: October 29, 1980
TIME: 1:30 p.m. - 3:30 p.m. LOCATION: Lincoln West

399

PRESENTATION OF PAPERS

CLINICAL PROBLEM SOLVING

MODERATOR: Geoff R. Norman, Ph.D.
McMaster University, Department of Family Medicine

Examination of the Effects of Structured Small Group Formats on Medical Students' Problem-Solving Performance
James M. Shumway, M.Ed., James F. Donohue, M.D.

Clinically Relevant Problem Solving Evaluation in Preclinical Medical Education: A Study of Alternative Approaches
Sarah A. Sprafka, Ph.D., Verda M. Scheifley, Ph.D.

Problem Solving Analysis: A Piagetian Study
James P. Hale, Ph.D.

DATE: October 29, 1980
TIME: 3:45 p.m. - 5:45 p.m. LOCATION: Georgetown East

PREDICTABILITY OF PREDICTIVE TECHNIQUES

MODERATOR: Roger Girard, Ph.D.
University of Southern California, School of Medicine

Path Analysis of Medical Student Performance Data
Charles P. Friedman, Ph.D.

Canonical Redundancy Analysis: A New Technique to Predict Performance
Beth Dawson-Saunders, Deane R. Doolen

Incremental Validity: The Old and New MCATs Compared
Charles P. Friedman, Ph.D., Carol Q. Porter, B.S.

The Relationship Between MCAT Subtest Scores and Performance in Medical School - The Impact of the Undergraduate Institution
Carter Zeleznik, Ph.D., Jon Veloski, M.A., Samuel Conly, Jr., M.D., Mohammadreza Hojat, M.A.

Restriction of Range and the Predictive Validity of the New Medical College Admission Test
Thomas J. Cullen, Ph.D., Charles W. Dohner, Ph.D., Percy D. Peckham, Ph.D., Werner E. Samson, Ph.D.

DATE: October 29, 1980
TIME: 3:45 p.m. - 5:45 p.m. LOCATION: Georgetown West

PRESENTATION OF PAPERS

DEVELOPMENT OF PATIENT ATTITUDES

MODERATOR: Miriam S. Willey, Ph.D.
Howard University, College of Medicine

Medical Students' Attitudes Towards Patient's Physical, Psychological and Health State Characteristics
C. Warner Johnson, M.D., Kaaren I. Hoffman, Ph.D.

Developing a Psychosocial Educational Program for Primary Care Physicians: Needs Assessment and Evaluation Baseline
Steven A. Cohen-Cole, M.D., M.A., John Boker, Ph.D., Julian Bird, M.R.C.P., M.R.C. Psych., Arthur Freeman, M.D.

Fostering Emotional Defensiveness in Intensive Care Unit Residents
Joel Frader, M.D., Charles Bosk, Ph.D., Ellen Prince, Ph.D.

DATE: October 29, 1980
TIME: 3:45 p.m. - 5:45 p.m.   LOCATION: Lincoln West

401

PRESENTATION OF SYMPOSIA

TOPIC: Explicit Definitions of Competence for Graduate Medical Education: What, How and So What?

Chairman: Philip G. Bashook, Ed.D.

Participants: Jerry D. Gates, Ph.D.
John S. Lloyd, Ph.D.

Discussants: Fredric D. Burg, M.D.
Leslie J. Sandlow, M.D.

DATE: October 29, 1980
TIME: 9:00 a.m. - 10:45 a.m. LOCATION: State

TOPIC: Cost Awareness Education and Practice of Medicine: Some National and International Perspectives

Chairman: Jack L. Mulligan, M.D.

Organizer: Mohan L. Garg, Sc.D.

Participants: Evert Reerink, M.D., Ph.D.
George A. Schlichte, Ph.D.
W. M. Kleinberg, M.D.
Mohan L. Garg, Sc.D.
W. A. Gliebe, M.A.

Discussants: Joseph Gonnella, M.D.
William Sodeman, Sr.
Mohan L. Garg, Sc.D.

DATE: October 29, 1980
TIME: 9:00 a.m. - 10:45 a.m. LOCATION: Military

TOPIC: Educational Models in Primary Care

Chairman: Edward A. Wolfson, M.D., M.P.H.

Organizer: Lawrence P. Tremonti, M.D., F.A.C.P.

Participants: Daniel L. Moser, M.S.
Craig Booher, M.D.
C. Kent Smith, M.D.
Robert H. Seller, M.D.

Discussants: W. Barry Biddle, Ph.D.
Lawrence P. Tremonti, M.D., F.A.C.P.

DATE: October 29, 1980
TIME: 9:00 a.m. - 10:45 a.m. 402 LOCATION: Hemisphere

PRESENTATION OF SYMPOSIA

TOPIC: Medical Student Education for Rural Practice: Influence of Curriculum and Learning Site

Chairman: Arthur Kaufman, M.D.

Participants: Arthur Kaufman, M.D.
Paul T. Werner, M.D.
Tom Cullen, Ph.D.

Discussants: Ronald Richards, Ph.D.

DATE: October 29, 1980
TIME: 11:00 a.m. - 12:45 p.m. LOCATION: State

TOPIC: Alternative Approaches to Research on Clinical Reasoning

Chairman: Arthur S. Elstein, Ph.D.

Organizer: Michael M. Ravitch, Ph.D.

Participants: Michael M. Ravitch, Ph.D.
David B. Swanson, Ph.D.
Georges S. Bordage, M.D., M.S.

Discussants: Barbara McNeil, M.D., Ph.D.

DATE: October 29, 1980
TIME: 11:00 a.m. - 12:45 p.m. LOCATION: Military

TOPIC: Perspectives on the Roles of Offices of Medical Education in the 1980's

Chairmen: Emil Petrusa, Ph.D.
Rose Yunker, Ph.D.

Participants: Frank Schimpfhauser, Ph.D.
Charles Dohner, Ph.D.
T. Joseph Sheehan, Ph.D.
Harold G. Levine, M.P.A.
Stephen Abrahamson, Ph.D.

DATE: October 29, 1980
TIME: 11:00 a.m. - 12:45 p.m. LOCATION: Hemisphere

PRESENTATION OF SYMPOSIA

TOPIC: A Multi-Institutional Research Study on the Use of Simulation for Teaching and Evaluating Patient Examination Skills

Chairman: Abdul W. Sajid, Ed.D.

Participants: Michael S. Gordon, M.D.
Joan W. Mayer, M.D.
Gordon A. Ewy, M.D.
Alan D. Forker, M.D.
Abdul W. Sajid, Ed.D.
Joel M. Felner, M.D.
Dorthea Juul
Robert A. Waugh, M.D.

Discussants: Howard S. Barrows, M.D.
Arthur S. Elstein, Ph.D.

DATE: October 29, 1980
TIME: 1:30 p.m. - 3:30 p.m. LOCATION: State

TOPIC: Recruitment and Selection of Physicians for Primary Care and Rural Practice -- Results from the Project TALENT Longitudinal Study

Chairman: Sandra R. Wilson, Ph.D.

Participants: Sandra R. Wilson, Ph.D.
Lauress L. Wise, II, Ph.D.
Frances B. Stancavage

Discussants: Lee Sechrest, Ph.D.
Paul Elliot, Ph.D.

DATE: October 29, 1980
TIME: 1:30 p.m. - 3:30 p.m. LOCATION: Hemisphere

PRESENTATION OF SYMPOSIA

TOPIC: Medical School Learning Environment: Prospects for New Research Directions

Chairman: Arthur Rothman, Ph.D.

Organizer: Sarah M. Dinham, Ph.D.

Participants: Sarah M. Dinham, Ph.D.
Christel A. Woodward, Ph.D.
T. Joseph Sheehan, Ph.D.

Critic: Edwin B. Hutchins, Ph.D.

DATE: October 29, 1980
TIME: 3:45 p.m. - 5:45 p.m. LOCATION: State

TOPIC: Learning During the Senior Clerkship

Chairman: Hugh M. Scott, M.D.

Participants: R. Kennedy Smiley, M.D.
Hugh M. Scott, M.D.
Vimla Patel

Discussants: W. Dale Dauphinee, M.D.

DATE: October 29, 1980
TIME: 3:45 p.m. - 5:45 p.m. LOCATION: Military

TOPIC: The *Standards* Developed by the Joint Committee on *Standards* for Educational Evaluation: Reflections on their Impact for Evaluation in Medical Education

Chairman: Hulda Grobman, Ed.D.

Participants: Hulda Grobman, Ed.D.
Phillip Bashook, Ed.D.
Vic Neufeld, M.D.

Discussants: Joseph S. Gonnella, M.D.

DATE: October 29, 1980
TIME: 3:45 p.m. - 5:45 p.m. LOCATION: Hemisphere

INDEX -- AUTHORS AND SYMPOSIA PARTICIPANTS

407