WORKSHOP ON THE NEEDS FOR RESEARCH
IN MEDICAL EDUCATION

Alexandria, 11 - 15 March 1974

EM/WRK.RES.MED.EDUC/4

ENGLISH ONLY

AN OVERVIEW OF APPLIED RESEARCH IN MEDICAL EDUCATION PROBLEMS, PRINCIPLES AND PRIORITIES*

by

Christine H. McGuire*
WHO Consultant

---

* Professor of Medical Education and Assistant Director for Research and Evaluation, Center for Educational Development, University of Illinois College of Medicine, Chicago, Illinois, U.S.A.

AN OVERVIEW OF APPLIED RESEARCH IN MEDICAL EDUCATION PROBLEMS, PRINCIPLES AND PRIORITIES*

Christine H. McGuire+

Over the past decade we have witnessed the development of critical shortages in health manpower in every region of the world. These shortages, though perhaps always to some degree extant, have obviously been intensified by rapidly increasing populations and sharply rising expectations across the globe. Less obviously, such shortages have also been exacerbated by the response of medical educators to concurrent changes in their scientific disciplines--a response that has been characterized by almost universal demands to increase the selectivity of entrants into the profession, to extend the period of required professional training and to further fragment that training in proliferating groups of loosely related sub-specialties. Clearly, this response has had the consequence, however unintentional, of shortening the average lifetime of professional service, of reducing the variety of health problems a highly specialized professional feels competent to manage and, hence, of restricting the supply of available health manpower at the very time that demands for such manpower are escalating.

It is not surprising therefore that most nations have responded by dedicating more and more resources to the expansion of educational facilities in the health professions. What is surprising is that, for the most part, this expansion has been designed simply to increase the annual output of physicians trained in the conventional image--an image that has already been recognized throughout the world as inappropriate for meeting the diverse health care needs of the total population. What was needed was a critical and imaginative application of research from educational and other behavioral sciences in the design of radically different programs to train new kinds of health professionals and to educate both old and new more efficiently and effectively. What was done was to expand existing institutions and to create new ones of the same type. If experience is any guide, it is more than likely that these increased facilities will be unable to produce physicians in the needed numbers, and that those which are produced will, like their predecessors, congregate in locations and in specialties where they will make the relatively smallest net contribution to meeting the health care needs of the country. In

*Prepared for discussion at the Workshop on the Needs for Research in Medical Education, Alexandria, 11-15 March, 1974.

+Professor of Medical Education and Assistant Director for Research and Evaluation, Center for Educational Development, University of Illinois College of Medicine, Chicago, Illinois, U.S.A.

short, what we are seeing in most places is the uncritical perpetuation of a system of medical education that has had demonstrable shortcomings everywhere it has been tried--shortcomings that have left practically every country in the world suffering not only from shortages of physicians, but also from their maldistribution with respect both to geography and to specialty.

This is not to deny that there have been some departures from the conventional pattern of 1-2 years pre-medicine, followed by 1-2 years of basic sciences, 2-3 years of clinical sciences and 1-5 years of specialty and sub-specialty training--a pattern which is now essentially worldwide. But, for the most part, modifications in the conventional program have been limited to such minor adjustments as adding a year here or there, combining pre-medicine with basic sciences or basic sciences with clinical experience to accelerate by a year or two, reducing hours in some disciplines and adding them in others, developing a somewhat more interdisciplinary approach in some areas, stressing somewhat more a comprehensive approach to patient care, introducing a few new subjects (e.g., genetics, human sexuality), replacing some requirements with electives, changing grading practices from the use of numbers to letters or even to words (e.g., pass, fail, honors), and supplementing traditional instructional modalities with small group and/or self-instructional techniques. Modifications such as these are all too reminiscent of Nero's "fiddling"; they do not represent basic reforms guided by systematic application of the findings from current research in medical education. Nor have the few totally innovative programs of physician education introduced in recent years been systematically directed by the insights obtained from this applied research. Even less have the effects of any of these changes been monitored by the utilization of techniques that are now readily available for gathering evidence that would facilitate more rational decision-making.

I have dwelt at this length on the patterns that have characterized our response to the needs for expanded health manpower because I believe that careful consideration of the research techniques and findings now available would have resulted in very significant modifications in the nature of education both in medicine and in other health professions. It is not yet too late to utilize these methods and results to inform the decisions that are still to be made. To this end I should like to call to your attention some of the newer techniques that are being increasingly employed to collect data and some of the implications of those data in four areas of research in medical education which I regard as critically important to you in your role as decision-makers in medical education:

Research on the goals of medical education

Research on the characteristics of medical students

Research on the setting for learning, curricula and instructional methodologies

Research on methods of evaluating professional competence

Research on the Goals of Medical Education

Basic research on learning indicates that students learn more, learn it more efficiently and retain it longer, when the objectives of an educational program are clear, when they are perceived by students as relevant to their own interests and motivations, and when they are shared by both faculty and students. Research in higher education indicates that this finding is as true of students in professional schools as to students at lower educational levels. The problem for the medical educator consists in selecting from among many worthy goals those of highest priority and then making these expectations explicit to students.

There are essentially two ways of making this selection: One consists in relying on the judgment of experts to determine what a neophyte in the profession ought to know and ought to be able to do. In the past we have relied almost exclusively on this method for determining the goals of medical education. As a result we find the curriculum crammed with an ever burgeoning quantity of new and highly specialized knowledge which the student perceives as irrelevant to his own goals and which, in fact, may be of little value to other than the super sub-specialist. Certainly expert opinion is an important source of information about the knowledge and skills which trainees should be able to demonstrate, but it is also possible to make this decision on the basis of scientific evidence about what competent physicians need to know and need to be able to do in order to deliver responsible patient care. A number of procedures have now been developed for collecting such data to provide an empirical basis for developing a behavioral description of the essential components of professional competence to guide a faculty in setting their goals and designing their curricula. Three of these may be of special interest to this group: the critical incident technique, the method of task analysis and the analysis of epidemiological data.

The Critical Incident Technique. This method consists in collecting comprehensive data about specific behaviors that characterize professional effectiveness and ineffectiveness and using these data to make an objective, empirical determination of the essential performance requirements of the profession. This technique is an outgrowth of studies in aviation psychology made in the United States during World War II. In that programme it was found that in reporting the reasons for eliminating a trainee, pilot instructors

and check pilots frequently offered such cliches and stereotypes as "lack of inherent flying ability," "poor judgment" or unsuitable temperament." In an effort to determine the specific qualifications of personnel that contributed to success or failure, combat veterans were asked to report incidents observed by them that involved behavior which was especially helpful or especially inadequate in accomplishing the assigned mission. This request concluded with the statement: "Describe the officer's action. What did he do?" The several thousand incidents submitted in response to this inquiry were analyzed and categorized to provide a relatively objective and concrete description of the "critical requirements" of combat leadership.

In applying this method to the medical profession, several thousand incidents describing observations of especially effective or especially ineffective colleague behavior are collected from several hundred physicians representing various age groups, geographic areas, types of affiliations and specialty interests. For example, in a critical incident study of intern and resident performance (i.e., of the general, undifferentiated physician) commissioned by the U.S. National Board of Medical Examiners, the American Institutes of Research who conducted the study collected over 3,000 incidents from physicians across the country. The incidents submitted involved all areas of behavior: cognitive, affective and psychomotor. They identified, for example, such general requisites of competence as "Skill in Gathering Clinical Information," i.e., in taking a competent history and in performing an adequate physical examination, or "Skill in Using Special Diagnostic Methods," i.e., in ordering and interpreting x-rays, biopsy specimens and the like, "Skill in Relating to the Patient and in Gaining His Cooperation in a Plan of Management." In an analogous study conducted by our own Center of the critical performance requirements in orthopaedic surgery, over 1,700 incidents were collected from over 1,000 orthopaedic surgeons representing various practice settings and subspecialty interests. An empirical classification defining 94 critical performance requirements, grouped into 9 major categories of competence, was derived from the incidents. This operational definition of the essential components of competence could then be employed to determine the goals of specialty training, the design of programs for their achievement and the criteria and methodology for their evaluation. If our educational planning were regularly directed by such operationally defined, empirically derived goals, I dare say our educational programs would look substantially different.

Task Analysis. A second method of determining the essential components of professional competence which should define the goals of medical education consists in detailed task analysis of what physicians in various

practice settings actually do, and in deriving from that statement of tasks a statement of the requisite knowledge and skills which they must have in order to perform these tasks competently. Such a task analysis may be based on careful, systematic observations of a representative sample of physicians in different practice settings, or it may be based on diary studies from the daily logs of a representative sample of physicians who report in minute detail the way in which they spend their professional days over a specified period of time, or on some combination of these two approaches.

Wherever this method has been employed, the results have been most enlightening. For example, in a limited pilot study of pediatricians in a typical small U.S. city, researchers found that all the physicians had different but consistent patterns for taking a history and performing a physical examination. Of the 481 patient visits observed, 222 were well children; an average of 10.2 minutes were spent with these children (range: 7.5 minutes to 13.6 minutes) in contrast with an average of 8.1 minutes spent with ill children (range: 7.4 minutes to 10 minutes). Of the 259 ill children, 104 (i.e., 40%) were diagnosed as having an infection of the upper respiratory tract, 15 had chronic illnesses and 5 had potentially life threatening diseases. For the total group of 481, optic fundi were examined only 9 times and rectals were performed in only 6 cases; 2 physicians did not percuss the lung fields for any patient. The greatest amount of time was spent in discussion of nutrition and child development. The single most frequent topic on which advice was rendered in well-child care concerned toilet training. The authors of this study concluded, "Few aspects of well-child care appear to require the skill of a physician...the question is also raised as to whether current training programs are aggravating the physician manpower shortage by overtraining in relation to community health needs."*

This is a question that I believe could be reiterated in every specialty in every country; only task analysis or comparable empirical sources will give us the answer.

Epidemiological Studies. One of the most interesting of the newer approaches to the use of such sources in determining the goals of medical education consists in combining three arbitrarily weighted factors--disease incidence, individual disability and social disruption to define priorities in health care needs and, hence, in educational effort. As

*Bergman, A., Probstfield, J. and Wedgewood, R. Performance Analysis in Pediatric Practice: Preliminary Report. Journal of Medical Education, Volume 42:262 (1967).

initially developed by Dr. John W. Williamson* the three factors are computed as follows: disease incidence consists of a simple tabulation of the frequency of the disease (e.g., pneumonia) or other medical condition (e.g., pregnancy) in the target population. Individual disability involves a determination of the extent of patient impairment or risk associated with a given medical condition; an Individual Disability Weight (IDW) is calculated for each condition from three elements: the average length of hospital stay, mortality rates and complication rates. Social disruption represents an estimate of the impairment that would be produced by a given disease or condition in the larger social group of which the patient is a member; it is based on such factors as cost of illness, age of patient and number of dependents, socio-economic standing and the like. For each discharged patient a Total Priority Weight (TPW) is calculated combining these elements. This Total Priority Weight is then arbitrarily apportioned among patient diagnoses. Finally, a cumulative total for each diagnosis is calculated from the total patient sample. The resultant ranking represents a quantitative estimate of priority or health care needs for the population at risk.

It is clear that even with unlimited resources not all of these needs can be met in the present state of our knowledge. The next step therefore consists in defining the area of total health care needs which can be met, given our present understanding of disease and our present resources for therapeutic interventions. This area therefore defines the target area for application of professional understandings and skills and helps to define educational priorities. The goals of medical education for the basic undifferentiated physician can therefore be defined as encompassing those areas of health care needs which cause the greatest total preventable disability--i.e., those which cause the greatest total disruption that could be reduced or minimized by early diagnosis and appropriate intervention.

In his early studies using this methodology to review hospital practice in two large community hospitals in widely separated metropolitan areas in the United States, Dr. Williamson found that pregnancy, involving uncomplicated delivery, ranked first or second in priority in both hospitals, that cerebral vascular accidents ranked among the first five diagnostic categories in both hospitals and that fractures of the lower extremity ranked among the first five in one hospital. I mention

*Williams, J. et al. Journal of American Medical Association, Volume 201:938 (1967) and Volume 204:303 (1968).

these particular conditions because in our own educational institutions there is a general tendency to reduce the amount of clinical instruction for the general medical student in some of these areas. For example, instruction in orthopaedic surgery is now an elective in our institution, despite the fact that trauma in general accounts for a very significant proportion of the total preventable disability in our society.

While the study I have reported above was limited to hospital practice, the same methodology could readily be applied to any level of health practice. Indeed, Dr. Williamson's current research involves just such an extension. Secondly, while the findings from such epidemiological studies and the particular weights to be assigned such factors as individual disability and social disruption will, of course, vary markedly in different parts of the world, the approach is clearly applicable to any population for whom health manpower are being trained. But, it is our suspicion that in all parts of the world, utilization of such data will modify the goals and priorities of our educational institutions and the emphases in our medical curricula, by focusing far greater attention on ambulatory medicine and on more common causes of disability than are frequently seen in our teaching hospitals.

Implications of Applied Research on Goals and Priorities

In the foregoing brief descriptions of methodology, I have tried to suggest that the means are now at hand for supplementing expert judgment with data derived from empirical studies to assist us in defining the roles and, hence, the requisite competencies of the graduates of our programs. I have no doubt that, should such studies be generally undertaken and should their findings be optimally employed in developing explicit goals and objectives of education in the health professions, we would see revolutionary changes in the kinds of health professionals produced, and in their education programs. Furthermore, I am convinced that such changes would have far greater impact in meeting the health care needs of our populations than the simple expansion of educational facilities of the conventional type.

Research on Student Characteristics

In contrast with the relatively limited and only recent attention that has been given to research on the goals of medical education, student characteristics have long been an object of intensive research and study, particularly as these characteristics relate to the problems of selection and attrition.

Research on Cognitive Characteristics

Predictive studies forecasting the probabilities of success for students with different entering characteristics represent the oldest and perhaps the best established area of research in this field. For the most part these studies have focused on using prior academic performance (with or without supplementary data from tests of scholastic aptitude) to predict success in medical school. In general, the criterion of "success" has been limited to academic record in medical school. It is not surprising that most of these studies reveal that prior academic performance, when appropriately weighted to take account of variations in the standards of different preparatory schools, is the best predictor of medical school grades. Of greater significance is the repeated finding that the correlation between previous academic record or aptitude test scores and academic standing in medical school rarely exceeds .5-.6; even this modest figure is progressively diminished as students proceed from the basic sciences to the clinical disciplines. Most important of all it has repeatedly been shown that class standing in medical school has no significant correlation with any criterion of physician performance. In short, we have been able to predict who will make good grades in the early years of medical school; we have been less successful in predicting who will make good grades in clinical years and we have met with essentially zero success in predicting who, from a large pool of qualified applicants, will make good physicians.

Research on Interests, Attitudes, Values and Personality.

As a result in some areas of the world, research on student selection is focusing less on the intellectual and academic attributes of the applicant population and focusing very much more on the personality, values, and interests of that population. This shift in emphasis has been accelerated by the conviction that, in order to have any significant impact on the geographic and specialty distribution of health manpower, we will need not only to redefine the goals of medical education but also in our selection of students to look more carefully at the attitudinal and other non-cognitive variables which influence career choices. Increasingly, such data are being sought for three purposes: to assist admissions committees in selecting students who are most likely to become primary health care providers; to furnish curriculum makers with concrete information about occupational, vocational and educational aspirations of students that will aid them in planning curricula to meet varied career goals; and to assist the teacher in implementing an instructional strategy that takes full advantage of what everyone already knows, namely, that students of

all ages learn more, learn it faster and learn it at less cost to the teacher if the latter can capitalize on the interests of the former.

In gathering specific information about the nature of student goals and interests we have relied primarily on the learner to tell us through his responses to the arsenal of ever more detailed questionnaires, activities indices, interest inventories and the like with which we bombard him. Research on this aspect of student need has been dedicated on the one hand to making these instruments both more subtle and more comprehensive, and on the other to developing more sophisticated analyses of the relationships between patterns of student responses and success in a career, to the end that data about conscious interests and needs can be used with ever greater confidence in student selection and career counseling.

However, it is my own personal judgment that research results to date do not justify heavy reliance on interest and need assessment for purposes of making decisions about individuals except perhaps at the extremes. Rather the most important implications for curriculum planners, teachers and counselors lies in the repeated finding that there is much more diversity in the life styles and in the patterns of interest and values within professional groups, and much more individual variation in career goals within student groups than we have been led to believe. It is time that our colleges of medicine were organized not only to recognize, but also to respond to and to take advantage of this rich resource in fulfilling our responsibility of educating health personnel to meet pressing societal needs.

Research on Personality in Relation to Learning Environment.

Related to this growing interest in a study of personality has been a concurrent upsurge of interest in looking at variations in student attitudes, values and personality as placing variable demands on the educational environment. In this work student characteristics are seen as unconscious demands on the environment and environment is seen as presenting an assortment of expectations and activities, pressures and rewards, facilities and people, all of which require some kind of adaptive response from the student. In contemporary personality theory student performance is seen as a function of the congruence between the student's unconscious needs and the potentially conflictful demands of the environment.

Research in this field has been dedicated primarily to the purpose of testing this hypothesis and a special methodology has grown up around it. Very briefly, this methodology consists in administering to groups of students a personality inventory of the "needs-assessment" type and a

corresponding "environment inventory." In early investigations the two inventories were usually designed to be entirely parallel insofar as a priori reasoning could make them so. For example, if the personality inventory included a series of statements to identify high need for structure, emotional support, achievement or some other satisfaction, the environment inventory administered to the same students would then include a series of statements to determine whether the environment is capable of responding to these same needs, or whether it is such as to create pressures of a conflicting nature. However, in his 1969 summary of the research based on this model Pace, who was one of the early investigators, concludes that, in general, the expected relationship between personality needs and environmental demands represented by the intended parallelism between personality and environment inventories, "has not been empirically demonstrated as fully as had been hoped."*

What has been documented is that educational environments do differ greatly, and what happens to students does depend in some measure on the particular character of the environment, but the influence of different patterns of environmental press on different patterns of personality need is exceedingly complex and its implications for educational planning and decision-making are far from clear. Indeed, Pace himself observes that "Many personality traits may or may not be relevant to objectives of higher education and the appropriateness of their use as criterion measures in studies of college impact is at least open to question."**

Even if this question were to be resolved, it is my own personal view that a far more vexing one presents itself. In professional education, specifically medical education, how shall we structure the environment to deal with personality needs which may be of negative value in the profession students expect to enter. For example--can we afford to create a medical school environment that reinforces a high need for structure when it is our task to educate students for a profession that requires lifetime learning in an independent, unstructured setting? Can we afford to create a medical school environment that reinforces a high need for certainty when it is our task to educate students for a profession that requires each to live with a high level of uncertainty? Perhaps the real question is--to what extent can we utilize findings from research on the relation between environmental press and personality needs to design a climate for learning that maximizes

*College Environments, In Encyclopedia of Educational Research. 4th Edition, Robert Ebel, ed. American Educational Research Association, 1970, p. 170.

**Ibid. p. 172.

the probability of accomplishing our institutional goals? Certainly, even at its present level of development research on need and environment, when considered in relation to contemporary theories of learning, has major implications for educational decision-making.

Research on Student Needs for Purposes of Diagnosis and Counseling

Closely associated with the investigation of attitudinal and personality variables has been an increased interest in research on methods of diagnosing student achievement in a manner which identifies any potential obstacles in the student's path toward the required level of mastery and which furnishes the basis for continuing counseling and educational prescription to him. Educators have long known student achievement in any area falls into a normal distribution curve. They have also repeatedly demonstrated that these individual differences in both level of attainment and rate of learning new material are exacerbated over time in the conventional, fixed, rigidly scheduled curriculum characteristic of most of our educational programs. Current research suggests, however, that the introduction of a program of systematic, diagnostic testing which is used to generate specific educational prescriptions can lead to a significant reduction in the range of individual differences in both ultimate achievement and in the time required to reach a given level of competence.

These studies suggest (1) that virtually all students of normal intelligence in any class are capable of achieving a high level of mastery; and (2) that if ample time is made available to each student to achieve mastery of the introductory units of instruction, individual differences in the amount of time required to reach that level are progressively diminished for subsequent units of instruction. While these effects are most pronounced in courses of instruction that require sequential organization of concepts and principles--e.g., algebra--they seem to occur even in courses of instruction in which no such organization is necessarily apparent--e.g., medicine.

In my view this finding has more profound implications for the educational decision-maker than any other finding of contemporary educational research. For, if it can be demonstrated in medicine, it implies nothing less than the total reorganization of our educational system and its associated instructional strategies, to be replaced by a completely individualized program based on scientific educational diagnosis and prescription analogous to the process of individual diagnosis and prescription we follow in clinical medicine. In such a system any implication of present rigidities will need to be eschewed with religious fervor.

If this reform were to occur, the responsibility of the educational decision-maker will become what it must be: first, to define the nature of

the tasks to be learned; second, to document the characteristics of the learners along the several dimensions discussed above; third, to create those conditions which maximize the probability that the required learning will take place. For, once the task has been specified and the characteristics of the learners identified, the conditions under which learning is most likely to occur can be clearly set forth.

Research on the Setting for Learning, Curricula and Instructional Methodologies

Fundamentally, the goals of education are not unlike the goals of therapy; i.e., to facilitate maximum functioning of the organism. Hence, the methodology for evaluating educational interventions is in essence analogous to that employed in evaluating any other therapeutic agent, and the pitfalls of research in this area are parallel to those in any field of clinical research. The three approaches to research on educational programs that are currently most widely employed and that appear to hold the greatest promise can be distinguished in terms of the primary object toward which data collection efforts are directed: student perceptions about the program, the nature of the educational process and the quality of the educational product. Each approach has its analogue in clinical research and, like the latter, its characteristic data type and source.

Figure 1: Approaches to Research on Educational Programs

Approach	Data Source	Clinical Equivalent
Assessment of perceptions	Reports of student reactions	Evaluation of patient's subjective feelings
Assessment of process	Observations of student-teacher interaction	Evaluation of therapeutic rationale in terms of pathophysiologic principles
Assessment of product, outcome or effect	Tests of student performance	Evaluation of therapeutic efficacy

Studies of Student Perceptions

This approach is designed to answer the question: "How did the students feel about the quality of the teacher, the teaching and the program?" Formal or informal student reactions and/or ratings constitute the data for answering the question. Typically these data are collected by interview and/or questionnaire that may range from highly structured to completely open-ended inquiries. Some institutions employ a standard rating form in which the student records his evaluation of some aspect of the program, ranking it in relation to others. Some institutions use questionnaires in which the student merely indicates what he liked or disliked, what he found valuable or useless, what experiences he would prefer to have expanded or reduced and the intensity of these feelings. In some instances these rating forms and questionnaires may be combined with a third type of instrument in which the student is asked about the specific feelings or actions which were engendered in him by some aspect of a program. In general, such questions are selected on the basis of contemporary learning theory and represent those aspects of the student-teacher relationship which are thought to influence the efficiency and effectiveness of learning. For example, a student may be asked, "Did you feel free to ask the instructor a question about something you did not understand?" "How often were you stimulated to think about additional applications of concepts and principles the instructor was discussing?" "How often did you go to the library to follow up on an idea that interested you?" and the like.

Provided the data are properly interpreted as simple statements of consumer satisfactions and dissatisfactions, applied research on student perceptions can be extremely valuable for two purposes: (1) to assist in the diagnosis of some of the causes of inadequate achievement and (2) to contribute to the identification and creation of conditions that will enhance motivational and other affective responses that facilitate learning. However, such data cannot be used to judge the educational efficacy of a program.

Studies of the Educational Process

In contrast with the assessment of perceptions which deliberately seek opinions about the merits of various elements in the educational program, process assessment is designed to determine what actually occurred in a particular program of instruction. The basic question which it attempts to answer is: "What is the nature and quality of the setting and of the communication between student and teacher that took place in that setting?" Data in response to this question are obtained either from systematic, impartial, skilled observation or from student responses to objective inventories identifying particular elements in the learning environment and asking the

student to indicate the extent to which each of these elements existed. The elements which are identified for observation or for student description are those which are thought to be important in facilitating or inhibiting achievement of the desired outcomes. For example, the students may be asked "How often did the instructor actually observe you while you were examining a patient?" "How often did the instructor provide feedback to you on the accuracy of your performance?" "How much opportunity did you have to practice a particular skill?"

Unfortunately, repeated experimental studies of the relative effectiveness of varied instructional methodologies--lecture, small group discussion, programed instruction, self-instructional materials and the like--have produced conflicting evidence and inconclusive results. Certainly, it may have been naive to expect any definitive outcome from experiments that on reflection appear to represent a simplistic search for the educational panacea--an all purpose miracle method, a pursuit that has proved no less futile than the quest for that medical chimera--the all purpose miracle drug. Rather than a "final answer," what has emerged from research of this type is clear evidence (1) that students learn in many different ways and at very different rates, (2) that the appropriate methodology depends on the nature of the educational objectives sought, and (3) that, whatever the instructional methodology utilized, the greatest learning takes place when students are actively involved in the learning process and when the material to be learned has the greatest apparent relevance to the students' own objectives. Consequently, it follows that the most effective program is one which provides genuine opportunity for these individual differences in learning style and learning speed, and in which the specific materials and instructional techniques have been chosen with a view to particular objectives that have been made explicit to both faculty and students.

Studies of the Educational Product

Whether a particular set of materials, techniques and conditions for learning is, in fact, optimal can be determined only by examining their consequences. It is to this issue that studies of educational product or outcome are addressed. All product assessment is designed to answer the question. "To what extent is the educational (therapeutic) agent effective in accomplishing the educational (therapeutic) goals?" If it is true that teaching is undertaken mainly for the purpose of producing student learning, it follows that program effectiveness can ultimately be assayed only in terms of the extent to which this intended outcome is achieved, and program efficiency can be judged only after the costs are documented and charged against the changes in student knowledge, judgment, skills, habits and attitudes which are realized.

Two types of data are therefore indispensable in answering the question posed in this approach: that derived from comprehensive measures of relevant aspects of student performance, and that obtained from careful cost accounting of teacher and student time expended, other resources consumed and undesired side effects induced by the educational program.

The techniques used to assess performance range from conventional tests of information, through sophisticated simulations of clinical problems, to long-term, systematic observations of the learner in varied professional settings, as elaborated below in the discussion of research on the evaluation of professional competence. However, since the evidence is mounting that, given adequate time, personal instruction and self-instructional resources, most anyone of normal intelligence and emotional stability can master any element among the objectives of higher education, no program evaluation is complete in the absence of cost data, and of cost/benefit calculations. In the final analysis therefore program evaluation necessarily entails economic considerations in determining whom a country can afford to educate and at what level of mastery, i.e., at what cost in resources. Unfortunately, at present the only usable data tend to be those concerned with the direct money costs attributable to the construction and maintenance of specialized facilities (e.g., lecture halls, student laboratories) and materials (educational films, slides, self-instructional programs). In the absence of hard data to the contrary, it is difficult to escape the suspicion that, because their costs are hidden, we continue to cling to conventional types of educational programs and to traditional methods that are not only ineffectual but may also be inexcusably uneconomic since these methods rely so heavily on the most costly resources of all--professional time and patient disability.

Research on the Evaluation of Professional Competence

We have now come full circle. This paper was introduced with a discussion of research on the goals of medical education; it will conclude with a discussion of research on the outcomes of medical education. Research on techniques of improving methods of evaluating professional competence requires first that the goals of the educational program be clearly specified, that they be defined in terms of behavioral changes that are to be brought about in the learner, that test situations be designed to sample these behaviors and that these test situations be administered to the learner to determine whether he is able to perform in the desired manner and at the prescribed level of competence. Research in this field has taken two forms: First, careful, systematic analysis of existing techniques of assessing student achievement and educational outcomes and second, the development of new and more effective methods of evaluating those outcomes of particular importance.

Repeated analyses of conventional examinations indicate that the traditional methods have serious shortcomings. First, these studies have shown that virtually all such examinations, both oral and written, both objective and subjective, measure only a very limited and perhaps the least important aspect of competence: namely, the ability to recall, rapidly and under stress, isolated fragments of information. Even the most costly techniques of oral examination have repeatedly been demonstrated to involve little more than repetition of the information which the examiner wants to hear. Secondly, the more subjective of these examinations (the oral and the essay) have repeatedly been demonstrated to yield such unreliable results as to be, for all scientific purposes, essentially useless. For example, careful studies of the correlation between grades assigned by different examiners regularly reveal an unacceptable level of interrater agreement. Careful analyses of a series of oral examinations suggest that they are subject not only to unreliabilities due to differences in standards between different examiners, but are also subject to the unreliability of sampling only a variable few of the many questions that could be asked. Finally, most practical examinations suffer from both of these deficiencies: namely, such a low level of inter-observer agreement that the grades may vary significantly depending on the observer and secondly, such great variability in the examination conditions as to result in totally unstandardized samples of student behavior. No reputable scientist would base decisions on such poor instrumentation.

Recent research has therefore been directed toward the development of more relevant, valid and reliable methods of evaluating various aspects of competence. The methods that have been developed encompass new types of paper and pencil tests, unconventional oral and practical examinations, reliable record audits, objective observations of performa nce in actual hospital and clinical settings and, of special significance a fascinating variety of simulation techniques. Most of us are familiar with simulation technology as employed in the education of business executives, airline pilots and even astronauts (space flight simulators). However, despite the extensive literature on its value, simulation has been employed in only very limited areas of professional education. Perhaps such limited exploitation is in part due to the vision of astronomical costs conjured up by the word "simulator." Yet, reduced to its essence, simulation consists merely in placing an individual in a realistic setting where he is confronted by a problematic situation that requires his active participation in initiating and carrying through a sequence of inquiries, decisions and actions. The situation must be designed so that each of these activities triggers appropriate feedback which can be utilized for subsequent decisions about pending action, decisions which will in turn modify the problem in different ways depending on the unique configuration of reactions and interventions each person makes. In this fashion a simulation can be evolved through many stages until it is terminated when the individual reaches an acceptable resolution or is faced by disruptive alternatives brought about by his own decisions and actions.

Recently, this essence has been captured in various simulation modalities that are economically and technologically feasible to use in both the instruction and assessment of three critical components of physician competence: skill in interpreting clinical and laboratory data, judgment in patient management and skill in dealing effectively with patients and colleagues. In testing interpretive skills, we are increasingly utilizing simulated clinical and laboratory data by means of photographic reproductions, sound simulators, sound movies and videotapes, three-dimensional models which have been developed to accept varied pathology, and even automated robots that can be programed to present combinations of findings which can be modified in an almost infinite number of ways. In the assessment of problem solving skills, we are using written simulations of the clinical situation, computer assisted simulations, automated robots and even live simulations in which an actor, a housewife, another student, or almost anyone can be "programed," to simulate a patient in an interview setting. The student's skills in data gathering, in crisis management, triage, office and patient management and his skills in communicating with the patient and in getting the patient's cooperation in a plan of management can be objectively observed, and his deficiencies objectively documented. We have even undertaken to develop simulated interviews with colleagues to test the student's skill in dealing with referral and consultation requests or simulated interviews in which he may be required to demonstrate that he can communicate effectively with one or more members of the health team by giving instructions to a simulated nurse, or requesting assistance for his patient from a simulated dietitian or social worker, or making a presentation and responding to the reactions of individuals who have been programed to take the role of other members of the health team in a simulated staff conference.

The essence of this technique is that it tries to imitate in a carefully standardized, controlled situation the kind of decision-making process which the physician must go through in a real life situation and to require of him the variety of responses that the real life situation would require. These simulations are intended to imitate, not to duplicate, life; this is both their greatest strength and, in the minds of some, their greatest limitation.* Opinion aside, present research reveals the following advantages of simulation techniques in evaluating professional competence:

*It is important to recognize that simulation is not an appropriate method for teaching or for testing all asepcts of performance. For example, simple recall of factual information is more economically and directly measured by conventional techniques of objective testing. At the other extreme, professional habits can be assayed only by careful and repeated observation over a long period of time in diverse settings.

(1) Examinations composed of such problems appear to the student as far more relevant than typical oral or conventional multiple-choice tests; such perceived relevance is at the very least psychologically beneficial. (2) This perception of relevance can be achieved without being dependent on the accidents of nature and the flow of real problems available at the particular place where, and the specific moment in time when, a practical examination is to be given. (3) Simulation makes it possible to predetermine precisely the exact task which students are to be required to perform, to focus on the elements of primary concern and to eliminate irrelevant and confusing complexities that would contaminate the assessment. (4) Simulation enables an examining body to standardize the task for all examinees and to do so without subjecting one or a few patients to repeated harassment by large numbers of students. (5) By standardizing the tasks and focusing on the most significant aspects in each, it is possible to sample the student's performance on a representative group of problems within a reasonable time frame. (6) When the exact tasks that are to compose an examination are precisely defined and pre-selected, it is possible to develop carefully specified criteria for judging student performance, to train examiners in applying these criteria and, thus, to achieve an acceptable degree of interrater reliability in scoring the examination. (7) In contrast with reality, simulation offers all students the opportunity to assume full medical responsibility for the work-up and management of "patients" without any risk whatsoever to anyone! (8) Furthermore, in carefully developed simulations, a lifetime of chronic disease can be collapsed into a half-hour problem, at each stage of which the student can be required to demonstrate his judgment and can be provided with feedback about his interventions in a form which is more instructive than life itself usually yields. (9) Finally, this prompt, specific and unambiguous feedback, characteristic of well designed simulations, makes examinations composed of such problems a powerful tool for the enhancement of learning.

In short, the evidence from studies of the use of simulation at several levels of medical education and with a variety of types of groups suggests that when these techniques are properly exploited, simulation exercises give considerable promise of being an extraordinarily powerful tool not only for purposes of evaluating more relevant aspects of professional competence, but also for purposes of instruction in, and research on, that most complex objective of all--problem-solving and clinical judgment. In our own studies of undergraduate, graduate and practitioner performance using these newer assessment techniques, the findings strongly suggest that substantial numbers of our students leave our colleges of medicine without yet being able to apply the fund of knowledge available to them, to employ adequate problem-solving strategies, to follow a systematic approach in their own decision-making, to respond comprehensively to patient needs or to make regular and effective use of the paramedical, consultant and community resources at

their command. If these deficits do indeed exist in any significant percentage of our graduates, the implications for reform in curricula and in instructional strategies is clear. At the very least, research on student evaluation has provided us with a methodology for determining the nature and extent of performance deficiencies in any individual or in any class. Its limited exploitation in medical education is not basically a scientific or even an economic issue; it is a social-psychological one.

The Present Status of Applied Research in Medical Education

Like other research, educational research is motivated by three strong drives: first, there is the intellectual curiosity which we all share about "how things work"--i.e., the domain of basic research. Second, there is always a need to solve urgent practical problems--i.e., the arena of applied research. Finally, there is a never ceasing need to collect data that will assist the educational decision-maker to make more rational choices--i.e., the territory for institutional or agency monitoring and study. In considering research stimulated by these motives, it is interesting to note that in the past we have often tended to create a sharp dichotomy between basic research and applied or action research and have allowed invidious distinctions to prejudice us against the latter, as if research motivated by pure intellectual curiosity were somehow more valuable to the community than that motivated by urgent problems that must be attacked or by desperate needs to allocate scarce resources more efficiently. I think it clear from the foregoing that I, for one, reject the values implied in these traditional distinctions; indeed, I have found that research in health professions education is particularly challenging and stimulating for the educator precisely because there are urgent problems to be solved, there are decisions to be made and, fortunately, the professional ethic characteristic of the medical scientist leads the responsible policy maker in that area to ask for evidence and, by and large, to respond to the implications of the data supplied him.

In this kaleidoscopic summary of the status of applied research on the goals of medical education, the characteristics of students, the efficacy of programs and the evaluation of professional competence, I have tried to suggest that we now have reliable techniques for collecting data and a wealth of specific findings to inform our decision-making. We are at the point where we can make rational educational choices based on the same scientific criteria that guide our clinical judgments. The means are at hand for us to adapt the methodology and collect the data relevant to the needs and conditions in our respective countries. It is my view that in this effort, research on the goals of medical education should be given the highest priority to assure that our decisions about the education and evaluation of our students are such as to meet the most urgent health care needs of our people.

REFERENCES

On Goals of Medical Education

Flanagan, J. D., The Critical Incident Technique. Psychological Bulletin 5:327 (1954)

Fivars, G. and Gosnell, Nursing Evaluation: The Problem and the Process. New York, The MacMillan Co. (1966)

Williamson, J. W., Alexander, M., and Miller, G. E. Continuing Education and Patient Care Research. Journal of the American Medical Association 201:938 (1967)

Williamson, J. W., Alexander, M. and Miller, G. E. Priorities in Patient-Care Research and Continuing Medical Education. Journal of the American Medical Association 204:303 (1968)

On Student Characteristics

Bloch, J. H. ed. Mastery Learning: Theory and Practice. Chicago, Holt, Rinehart and Winston, (1971)

Crites, J. Interests. In Encyclopedia of Educational Research. Fourth Edition. Robert L. Ebel, ed. New York, MacMillan Co. (1969)

Gee, H. H. and Glaser, R. J. The Ecology of the Medical Student. Report of the Fifth AAMC Teaching Institute. Evanston, Illinois (now Washington, D.C.) Association of American Medical Colleges (1958)

Haley, H., D'Costa, A. and Schaefer, A. eds. Personality Measurement in Medical Education. Washington, D.C. Association of American Medical Colleges (1972)

McGuire, C. H. Diagnostic Examinations in Medical Education. Geneva, World Health Organization WHO/EDUC/72.151 (1972)

Pace, C. R. College Environments. In Encyclopedia of Educational Research. Fourth Edition. Robert L. Ebel, ed. New York, MacMillan Co. (1969)

On Evaluation of Professional Competence (General)

Bloom, B. S., Hastings, J. T. and Madaus, G. Handbook on Formative and Summative Evaluation of Student Learning. New York, McGraw Hill (1971)

Bloom, B. S. et al. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. New York, McKay (1956)

Charvat, J., McGuire, C. and Parsons, V. A Review of the Nature and Uses of Examinations in Medical Education. Geneva, World Health Organization, Public Health Papers No. 36 (1968)

Krathwohl, D. R., Bloom, B. S. and Masia, B. B. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook II: Affective Domain, New York, McKay (1964)

McGuire, C. A Process Approach to the Construction and Analysis of Medical Examinations. Journal of Medical Education 38:556 (1963)

McGuire, C. The Oral Examination as a Measure of Professional Competence. Journal of Medical Education 41:267 (1966)

On Evaluation of Professional Competence (Use of Simulation Technique in Medical Education)

Barrows, H. Simulated Patients. Springfield, Illinois, Charles C. Thomas (1972)

Harless, W. et al, CASE: A Computer-Aided Simulation of the Clinical Encounter. Journal of Medical Education 46:443 (1971)

Levine, H. and McGuire, C. The Validity and Reliability of Oral Examinations in Assessing Cognitive Skills in Medicine. Journal of Educational Measurement 7:63 (1970)

Levine, H. and McGuire, C. The Use of Role-Playing to Evaluate Affective Skills in Medicine. Journal of Medical Education 45:700 (1970)

McGuire, C. and Babbott, D. Simulation Technique in the Measurement of Problem-Solving Skills. Journal of Educational Measurement 4:1 (1967)

McGuire, C. and Solomon, L. ed. Clinical Simulations: Selected Problems in Patient Management. New York, Appleton-Century-Crofts and Co. (1971)

McGuire, C. and Solomon, L. and Bashook, P. Written Simulations: Their Construction, Scoring and Analysis (In Press)

McGuire, C. and Wezeman, F. Simulation in Instruction and Evaluation in Medicine. WHO/EDUC (In Press)

Penta, F. and Telder, T. Audiovisual Aids to Learning. WHO/EDUC Also in: Development of Educational Programmes for the Health Professions. Public Health Paper #52, Geneva, World Health Organization (1973)

On Programme Evaluation

McGuire, C. The Evaluation of Teachers and Teaching Effectiveness. WHO/ EDUC/73.165 (1973)

On Research Design (For the reader who desires a comprehensive, sophisticated discussion of the merits and limitations of alternative research designs, that reader will find these volumes an extraordinarily valuable resource in suggesting alternative solutions to the problems of research design in this area)

Gage, N. L., ed. Handbook of Research on Teaching. American Educational Research Association, Chicago, Rand McNally and Co. (1963)

Travers, Robert M. W. ed. Second Handbook of Research on Teaching. American Educational Research Association, Chicago, Rand McNally and Co. (1973)

Summary of Research in Medical Education

Miller, G. E. Medical Education. In Encyclopedia of Educational Research. Fourth Edition Robert L. Ebel, ed. New York, MacMillan Co. (1969)