DOCUMENT RESUME

ED 134 678 CE 008 413

AUTHOR Andrew, Barbara J.; And Others
TITLE National Program for the Evaluation of Primary Care. Physician's Assistants. Final Report.
INSTITUTION National Board of Medical Examiners, Philadelphia, Pa.
SPONS AGENCY Health Resources Administration (DHEW/PHS), Bethesda, Md. Div. of Associated Health Professions.
PUB DATE Jul 76
CONTRACT HRA-N01-AH-34031
NOTE 177p.; Parts of appendices will not be legible due to blurred trpe

EDRS PRICE MF-$0.83 HC-$10.03 Plus Postage.
DESCRIPTORS *Certification; *Criterion Referenced Tests; Evaluation Methods; Interpersonal Competence; Measurement Techniques; Medical Education; National Competency Tests; Occupational Tests; *Physicians Assistants; Predictive Ability (Testing); Primary Health Care; Simulation; Statistical Analysis; *Test Construction; *Test Validity

ABSTRACT

This report of the development and validation of a criterion-referenced examination for physician's assistants contains an overview of phase 1 project activities (1972-74) and more detailed description and data on phase 2 activities (1975-76). The overview describes the rationale for the development of the examination. (It was the first certifying examination for health care professionals that was based on task analyses of health care functions--data gathering, analysis and interpretation, and medical and health care strategies--as opposed to curricular content.) Administration of the first National Certifying Examination for Primary Health Care Physician's Assistants in 1973, the validation studies of that exam, and the administration and statistical analyses of the 1974 exam are summarized in the overview. Phase 2 activities are described in detail in three sections: (1) A study of audiovisual simulation techniques for assessing interpersonal skills, (2) a validation study of the 1974 certifying examination involving medical record audits, and (3) an analysis of the 1975 certifying examination. Appendixes A, B, and C contain the audiovisual simulation materials which include the sample paper-pencil simulation, the interaction observation form, and the patient protocol form. Appendixes D and E contain the audit validation study materials, which include the sample audit criteria set and the instruction manual and sample abstracting form for record reviewers. References are provided. (TA)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document Reproductions supplied by EDRS are the best that can be made from the original.

ED134678

NATIONAL BOARD OF MEDICAL EXAMINERS

NATIONAL PROGRAM FOR THE EVALUATION OF PRIMARY CARE PHYSICIAN'S ASSISTANTS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL REPORT

July, 1976

Contract Number: HRA-NO1-AH-34031

Submitted to the Department of Health, Education, and Welfare, Health Resources Administration, Bureau of Health Manpower, Division of Associated Health Professions

By the National Board of Medical Examiners

Barbara J. Andrew, Ph.D., Project Director
Vivian F. Erviti, Ph.D., Co-Director
Fred J. Dowaliby, Ph.D., Co-Director

CE 008 413

TABLE OF CONTENTS

OVERVIEW OF PHASE I ACTIVITIES

BACKGROUND ON EXAMINATION DEVELOPMENT

When work on this examination program was initiated in July, 1972, two primary goals were identified: (a) to specify the roles and responsibilities of the primary care physician's assistant, and (b) to identify evaluation techniques that could be used to assess competence in performing the functions of a physician's assistant. The development of examinations for health care professionals had long been based upon curricular content: the subject matter presented and the way in which its presentation was structured in the curriculum. However, because of the goals of this examination program, an approach was devised which had not previously been used for the development of national examinations for health care professionals. Instead of relying upon descriptions of the subject matter content within the various educational programs, a task inventory consisting of several hundred health care functions was designed. These functions were compiled from several task analysis studies of physician and non-physician activities. Once compiled, each health care function was classified under one of ten categories reflecting the various components of the clinical problem-solving and management process. These categories are listed below:

I. Data Gathering
   - A. History Taking and Patient Records
   - B. Physical Examination
   - C. Laboratory Tests and Investigative Procedures
   - D. Patient Monitoring

II. Analysis and Interpretation
   - A. Consultation and Referral
   - B. Diagnostic Acumen

III. Medical and Health Care Strategies
   - A. Emergency Procedures
   - B. Surgical and Technical Procedures
   - C. Management
   - D. Patient Counseling

The resulting task inventory was distributed to each member of the National Board Advisory Committee on Physician's Assistants whose membership included physicians involved in the development of this new category of health care personnel, physician's assistants, physicians who were training and utilizing physician's assistants, and nurses concerned with the nurse practitioner concept. These individuals were asked to consider each of the 900 health care functions included in the inventory, and to decide whether the health care function was one that a Type A primary care physician's assistant should definitely, probably, probably not, or definitely not be skilled in performing. Since the examination to be developed would be administered on a national basis and would be evaluating individuals trained in different types of educational programs, it was felt that the examination should be designed to measure those health care functions that a representative group of experts felt were ones the physician's assistant should definitely be skilled in performing. In order to identify the health care functions about which there was such a consensus, a frequency distribution of the twenty judgments made concerning each health care function was tabulated for all functions listed on the task inventory. Each of the four rating categories was given a numerical value on a scale of one to four a value of four being assigned to the "definitely" category, and a value of one to the "definitely not" category. In turn, each judgment made by a rater was given in numerical value on the basis of the column into which a check mark had been placed.

The arithmetic mean of these judgments was then computed for each health care function. Those functions receiving a mean value of 3.5 to 4.0 were considered as ones that, in the view of expert opinion, a Type A primary care physician's assistant should definitely be skilled in performing. Functions receiving a mean value of 3.40 to 3.49 were reviewed by this expert group at a meeting during which the results of the task inventory study were presented.

A review of the health care functions identified from this task inventory study suggested that not all functions were equivalent in terms of their importance to the proficiency of the primary care physician's assistant. Moreover, the number of functions was so large that it was evident that no examination program could attempt to sample adequately the knowledge and skills related to all of them. For these reasons, a priority study was conducted using the same experts who had participated in the task inventory study, plus four additional experts, all of whom were pediatricians.

The purpose of the priority study was to establish the relative importance of the several hundred health care functions, so that those receiving the highest rating in this regard could provide the primary focus for the development of the Certifying Examination for Primary Care Physician's Assistants.

Two dimensions were selected as the basis for determining the priority of each health care function: (1) the frequency with which the function might be carried out in a primary care practice, and (2) its "criticalness" to optimum health care delivery. In order to determine the priorities of the several hundred health care functions identified from the first phase of this process, a second task inventory was prepared using the categories previously described. The two dimensions which were the basis for arriving at these priority ratings (i.e., frequency and criticalness) were represented by two scales each consisting of four intervals. The anchor points of each scale were labeled "high" and "low". Each member of the group of expert judges was asked to consider the functions included on this second inventory, and to indicate how frequently he felt the task would be performed in a primary care practice and how critical it was to optimum health care delivery.

A frequency distribution of the judgments made regarding the frequency and criticalness of each health care function was tabulated. A mean value for each health care function on each dimension was calculated by assigning a numerical value of one to four to the intervals on each scale as was done in the first task inventory study.

The priority value assigned to each health care function was determined by using the following formula $P = \bar{f} + 2(\bar{c})$, where $\bar{f}$ was the mean frequency value and $\bar{c}$ the mean criticalness value. The criticalness value was weighted more heavily than the frequency value because it was recognized that while some functions are performed infrequently (e.g., closed chest cardiac massage), they often involve life-and-death implications when they are required as part of health care delivery. The use of the preceding formula resulted in a scale whose range was 3.00 to 12.00. Those functions receiving a priority of 8.00 to 12.00 were identified as high priority health care functions and were further analyzed in the following manner.

Since the validity of an examination as an assessment of proficiency depends upon its capacity to evaluate accurately the knowledge and skills required to carry out specific health care functions, test committees were appointed to analyze related health care functions and to identify the knowledge and skill components related to each. These knowledge and skill components were stated behaviorally to facilitate the selection of appropriate evaluation methodology and to serve as the performance criteria to be used in assessing the proficiency of physician's assistants. Table 1 presents some of the criteria specified for the following health care functions: history-taking; physical examination; wound care; suturing lacerations; taking an electrocardiogram; and applying a plaster case. (A more detailed description of this phase of examination development is provided in the Project Report 1973-74.)

Table 1

SAMPLE CRITERIA FOR SELECTED HEALTH CARE FUNCTIONS

Function	Criteria
History Taking	(a) Elicit pertinent positives and negatives related to chief complaint; (b) Elicit pertinent data from review of systems.
Physical Examination	(a) Recognize and name common skin lesions; (b) Recognize and describe abnormalities seen on funduscopic examination.
Wound Care	(a) Recognize normal signs associated with wound healing; (b) Identify lacerations requiring primary closure by suture.
Suturing Lacerations	(a) Select appropriate suture materials; (b) Effect a closure in which all dead space has been obliterated.
Electrocardiogram	(a) Identify the correct placement of all "V" leads; (b) Select appropriate courses of action to eliminate or minimize technical defects in an EKG tracing.
Cast Application	(a) Identify the correct position for applying a forearm cast; (b) Recognize errors made in case application.

ADMINISTRATION OF THE FIRST CERTIFYING EXAMINATION - 1973

The first National Certifying Examination for Primary Care Physician's Assistants was administered on December 12, 1973, to 880 candidates in 38 test centers across the country. Eligibility for the first examination was limited to graduates of primary care physician's assistant training programs that were approved by the AMA Council on Medical Education, funded by the Bureau of Health Resources Development, or, in the case of nurse practitioners, programs of at least four months' duration within a nationally accredited school of medicine or nursing that trained pediatric or family nurse practitioners.

Of the total number of physician's assistants eligible, 75 per cent registered for this examination, while almost 100 per cent of the eligible Medex did so. Of the nurse practitioners who were eligible for the examination, approximately 10 per cent registered.

In relation to the 880 candidates who took this examination, 62 per cent had received their training in physician's assistant programs, 29 per cent in Medex training programs, and 9 per cent in nurse practitioner programs. Table 2 summarizes the composition of the examinee group by type of training program.

Table 2

COMPOSITION OF EXAMINEE GROUP - 1973

Type of Training	No.	%
Physician's assistant	538	62
Medex	265	29
Nurse practitioner	77	9
Total	880	100

Biographic data collected on each examinee indicated that a significant majority had already completed their formal training prior to the examination, and had already acquired clinical experience as a physician's assistant. Table 3 summarizes the clinical experience of this examinee group. As illustrated in the table, 89 per cent of the examinees had already completed an educational program, and 81 per cent had already acquired postgraduate clinical experience prior to taking the Certifying Examination. Of those examinees who had already acquired experience as a physician's assistant, 88 per cent had had as much as two years of clinical experience, while 12 per cent had acquired more than two years of clinical experience as a physician's assistant.

In addition, 91 per cent of the total examinee group had acquired experience in health care delivery prior to their training as a physician's assistant. Of this group with prior experience in health care delivery, 86 per cent had been involved in direct patient contact (for example, as a nurse, military corpsman, or physical therapist), while 11 per cent had been involved in health care delivery in a technical capacity. These data indicate that the typical examinee for the 1973 Certifying Examination had already completed a formal educational program and had already acquired several months to several years of clinical experience as a physician's assistant. Moreover, prior to having been trained as a physician's assistant, the typical examinee had been involved in direct patient contact health care delivery for a period of from two to four years.

Table 3

EXPERIENCE OF EXAMINEE GROUP - 1973

Biographic Data	% - Total Group	
Completed an educational program	89	
Clinical experience since training:	81	
Up to 2 years clinical experience		(88)
More than 2 years clinical experience		(12)
Prior experience in health care delivery:	91	
Patient contact		(86)
Technical		(11)
Other		(3)

The 1973 examination program consisted of a one-day written examination divided into two sections. The first section contained multiple-choice and other objective format questions presented in printed and pictorial form. These materials were designed to assess the candidate's knowledge and skill in applying knowledge related to high priority health care functions that a primary care physician's assistant should be skilled in performing. Items on this section of the examination were classified under the following headings:

1. The identification and classification of physical findings;
2. Patient management;
3. Patient counseling and instruction;
4. Knowledge related to clinical procedures (e.g., wound care, fracture management, cardiopulmonary resuscitation, electrocardiograms).

The afternoon section of the examination involved a programmed testing technique in which the candidate was presented with a simulated clinical case and asked to make decisions regarding the appropriate diagnostic work-up and management of the patient as he would in an actual clinical setting. These patient management problems are designed to assess the candidate's skill in gathering pertinent information about patients and in making appropriate disposition decisions. Clinical cases were presented in both adult and pediatric medicine, and included emergency as well as non-emergency problems.

Statistical analysis of the examination indicated that it was reliable and moderately difficult for the group of examinees who took it. The reliability of the total examination was .89, which placed it within the range of reliabilities for other National Board examinations. The mean difficulty level for the multiple-choice question section of the examination equaled .64 which was also within the range encountered on other National Board examinations. The mean difficulty level of the patient management problems was calculated at .79 which also corresponds to the difficulty levels found on other examinations using patient management problems. These examination statistics are summarized in Table 4.

Table 4

EXAMINATION STATISTICS - 1973 PHYSICIAN'S ASSISTANT CERTIFYING EXAMINATION

Examination Component	Reliability	Range/Other NBME Examinations	Mean Difficulty	Range/Other NBME Examinations
Total	.89	(.88-.91)		
Multiple-choice questions			.64	(.60-.65)
Patient management problems			.79	(.75-.85)

Further analysis of examination performance in relation to biographic data indicated that examinees who had already completed a training program and had acquired clinical experience as a physician's assistant or nurse practitioner scored significantly higher on the examination than did examinees without such postgraduate clinical experience. This finding provides evidence of the construct validity of the examination since it appears to be measuring knowledge and skills that are relevant to practice and that increase with clinical experience.

Table 5 presents the intercorrelations of the various examination components. The correlation coefficient of .51 between the multiple-choice (MCQ) and patient management (PMP) sections of the examinations indicate that, although these components were assessing some overlapping areas of knowledge and skills, each was assessing aspects of competence that the other examination component was not. This correlation coefficient is similar to those encountered on other examinations for licensure and certification and adds further support to the tenet that these two different examination formats do assess different aspects of clinical competence.

Table 5

INTERCORRELATIONS OF 1973 EXAMINATION COMPONENTS

	MCQ	(PMP) Data Gathering	(PMP) Management/ Therapy	(PMP) Total
MCQ	—	.46	.45	.51
(PMP) Data Gathering		—	.36	.98
(PMP) Management/Therapy			—	.55
Total				—

Items on the patient management problems (PMP) were divided into two categories: (1) data gathering procedures (history, physical examination, and laboratory tests); and (2) management/ therapy decisions. The correlation between these two components of the patient management problems equaled .36, and indicates that items in each category are also assessing different aspects of clinical competence.

VALIDATION STUDIES OF THE 1973 CERTIFYING EXAMINATION

In order to investigate the validity of this examination program, a number of studies were conducted and analyses performed. The first evidence of construct validity was provided by an analysis of the examination itself in relation to certain biographic data that had been collected on all examinees. The fact that examinees who had already completed a training program and had acquired clinical experience as a physician's assistant or nurse practitioner scored significantly higher on the examination than did examinees without such postgraduate clinical experience suggests that the examination was measuring knowledge and skills relevant to clinical practice that increase with clinical experience. Since examinees with experience in patient contact health care delivery prior to training as a physician's assistant did not score significantly higher than individuals without such prior experience suggests that the examination was measuring aspects of competence that are specifically pertinent to the proficiency of a physician's assistant or nurse practitioner.

In addition to this internal evidence of validity, two external studies were also conducted, each focusing on somewhat different aspects of the validity of this examination program.

Having already found evidence to suggest that the examination was measuring knowledge and skills relevant to actual clinical practice, a study was conducted in order to determine whether or not the examination was measuring components of competence that could be attributed to the training process itself. If the Certifying Examination was measuring knowledge and skills that could be acquired only through a training program for physician's assistants, then one would expect examinees who sat as candidates for the 1973 Certifying Examination to score significantly higher than individuals who were just beginning a training program and who, presumably, had not yet acquired the same level of proficiency.

In order to investigate this aspect of the construct validity of the examination, the 1973 Certifying Examination was administered to groups of individuals who were just beginning training as physician's assistants, Medex, or nurse practitioners. A stratified random sample of 16 training programs was selected to be representative of the types of educational programs in which actual candidates for the 1973 examination had been trained. As each program began training of a new class, the 1973 Certifying Examination was administered to the group of trainees. The same procedures used in administration of the actual Certifying Examination were also used in administering the examination for validation purposes.

The validation study sample consisted of 357 examinees: 83% physician's assistants, 15% Medex, and 3% nurse practitioners. The actual candidate group in 1973 consisted of 880 examinees: 62% physician's assistants, 29% Medex, and 9% nurse practitioners. The validation study statistics, however, were generated by comparing the validation sample of 357 examinees with candidates for the 1973 Certifying Examination who had come from the same educational programs that participated in the validation studies. This actual candidate group numbered 528.

A statistical comparison of the performance of these two groups was carried out by performing t-tests on the total examination score as well as on scores obtained on the three components of the written examination. With respect to overall examination performance, the mean score of examinees from the validation study sample was 374, while a comparable sample of examinees from the actual 1973 examination obtained a mean score of 497. This difference between means is highly significant ($p < .001$) and indicates that individuals who had completed or were nearing completion of their formal educational program scored significantly higher than individuals who were just beginning training as a physician's assistant. Similar differences in performance were also observed on each of the three components of the written examination: multiple-choice questions, PMP data gathering, and PMP management/therapy. Thus, the 1973 Certifying Examination was able to distinguish between individuals with and without formal training as a physician's assistant. These statistics are summarized in Table 6 shown below.

Table 6

EXAMINATION SCORES OF INDIVIDUALS BEGINNING FORMAL TRAINING (UNTRAINED) AND CANDIDATES FOR THE 1973 CERTIFYING EXAMINATION (TRAINED)

Exam Component	Untrained	Trained	t	Significance
Composite	374	497	21.45	< .001
MCQ	341	503	22.52	< .001
PMP Data Gathering	377	493	16.45	< .001
PMP Management/Therapy	403	495	12.14	< .001

A third validation study was conducted in order to estimate the concurrent validity of the Certifying Examination. This study was designed to investigate the relationship between performance on the various components of the Certifying Examination and ratings of clinical competence as provided by program faculty. A clinical competence rating form was developed consisting of 40 statements each describing different aspects of the competency of physician's assistants. These statements described behavior related to: (a) history-taking, (b) physical examination, (c) laboratory tests and diagnostic procedures, (d) management/treatment, (e) medical records, and (f) interpersonal relations.

Individuals who had registered for the 1973 Certifying Examination and who were still in training or had recently completed training were identified. A clinical competence rating form for each of these individuals was mailed to the appropriate program director who was asked to identify a member of his faculty who was familiar with the clinical performance of each examination candidate. The faculty rater was asked to read each statement contained in the rating form and decide at what level of competence the candidate performed. A scale consisting of five intervals was used with the end points labeled "minimum competence" and "optimum competence."

Of the 179 rating scales mailed to program directors, 154 completed forms were returned, representing an 86 per cent return rate. Of this number, four were deleted from the sample since the registrants did not take the Physician's Assistant Certifying Examination. Thus, the subjects of this study were 150 registrants for whom a set of examination scores and ratings of clinical competence were available.

Since only a sampling of the total registrant pool was involved in the study, a preliminary analysis was performed to determine whether or not these subjects were representative of the registrant population. Accordingly, significance tests were performed between the study subjects and the remainder of the examinee population in terms of the distribution of performance on each major examination component. These comparisons consisted of t-tests between the means and F-tests between the variances of the two examinee groups. In all cases, these statistics were not significant, indicating no significant differences between the groups in terms of their performance on the major components of the examination. These results suggest that the study sample was representative of the population of examinees in terms of examination performance.

A factor analysis of the items contained on the rating scale was performed in order to identify clusters of related items and the dimensions of clinical competence each cluster appeared to be measuring. The factor analysis yielded the following three dimensions: (1) Data Gathering and Recording, (2) Interpersonal Skills, and (3) Clinical Judgment. The Data Gathering and Recording factor consisted of rating items assessing competence in taking histories, performing physical examinations, and recording patient data. The Interpersonal Skills factor contained rating items pertinent to interactions between the physician's assistant and the patient, the patient's family, and other members of the health team. The Clinical Judgment factor consisted of items pertinent to various aspects of patient management. Analysis of the examination and rating scale data was performed by correlating an individual's clinical competence rating on each of the three factors with the various examination components. In order to provide a more detailed analysis of the examination itself, items on the multiple-choice question portion were classified into one of the following four categories: (1) identification and classification of physical findings, (2) patient management, (3) knowledge of clinical procedures, and (4) interpersonal skills.

Pearson product-moment correlation coefficients were computed to determine the direction and magnitude of the simple correlations between examination components and rating factors. In those instances where an examination component correlated significantly with more than one rating factor, step-wise multiple regression analyses were performed to identify the nature and magnitude of these complex relationships. Evidence for the construct validity of the Certifying Examination would be provided by the extent to which significant positive correlations were observed between those examination components and rating factors where it would be logical to expect such a relationship. The results of these analyses are summarized in Tables 7 and 8.

Candidate performance on multiple-choice questions dealing with the identification and classification of physical findings correlated significantly with ratings of Data Gathering and Recording skills and Clinical Judgment. One would not expect performance on this examination component to correlate with ratings of Interpersonal Skills, and, in fact, such a relationship was not observed. Performance on multiple-choice questions related to patient management correlated significantly with ratings of Clinical Judgment, but not with ratings of Data Gathering and Recording or Interpersonal Skills, a pattern of correlations which is consistent with the logical constructs of these examination and rating scale components.

Similarly for patient counseling and instruction, examination performance on these multiple-choice questions correlated significantly with ratings of Interpersonal Skills and Clinical Judgment. One would not expect a significant relationship to ratings of Data Gathering and Recording skills, and none was observed. Performance on the PMP data gathering section (which included history-taking, physical examination, and selection of laboratory tests) correlated significantly with ratings of Data Gathering and Recording skills and Clinical Judgment. Since Interpersonal Skills were not measured on this component of the examination, one would not expect to find a significant relationship with ratings of Interpersonal Skills, and this was observed to be the case. Of the four examination components thus far described, all were found to correlate significantly with the rating factors to which they bore the most logical relationship and not with rating factors that did not seem pertinent to the nature of the examination component. To this extent, the pattern of significant correlation provides evidence for the convergent and discriminant validity of the Certifying Examination.

As will be noted in Table 7, multiple-choice questions related to the knowledge of clinical procedures did not correlate significantly with any of the three rating factors. This, too, is consistent with evidence for the discriminant validity of the Certifying Examination since none of the rating factors dealt with a candidate's level of competence in performing clinical procedures. Thus, the absence of significant correlation is consistent with the logical constructs of the examination components and rating factors. The one examination component for which evidence of discriminant validity was not observed was that component of the patient management problems that dealt with management and therapy decisions. One would have expected to observe a significant correlation between this examination component and ratings of clinical judgment. Although the magnitude of the correlation coefficient between PMP management/therapy and ratings of Clinical Judgment was greater than that observed for the other two rating factors, it did not reach statistical significance. A further analysis of the statements contained on the Clinical Judgment rating factor indicated that behavior related to the selection and sequencing of appropriate diagnostic tests and procedures made up a large number of the items included on this rating factor. However, the examination component labeled PMP management/therapy did not include those items on the patient management problems related to the selection and interpretation of diagnostic tests and procedures. Instead, these items were included as part of the PMP data gathering component. It is possible, therefore, that the lack of significant correlation was attributable to the incorrect classification of patient management problem items related to the selection of laboratory tests. Such a reclassification of these items has been performed for the 1974 Certifying Examination, and this analysis will again be performed in order to provide further data concerning the discriminant validity of this examination component. Table 8 summarizes the multiple correlation coefficients obtained for three of the six examination components. The magnitude of the multiple correlations is greater than that for the simple correlations because more than one rating factor is being used to predict performance on a single examination component.

Table 7

Simple Correlations Between Rating Factors
and Components of Examination Performance (N=124)

Examination Component	Rating Factor		
	Data Gathering and Recording	Interpersonal Skills	Clinical Judgment
Identification and Classification of Physical Findings	.18 *	-.02	.25 **
Patient Management	.08	.11	.21 *
Patient Counseling and Instruction	.13	.23 *	.23 *
Knowledge of Clinical Procedures ***	.04	.04	.16
PMP Data Gathering	.29 **	.12	.25 **
PMP Management/Therapy	.06	.07	.11

* Significant beyond the .05 level
** Significant beyond the .01 level
*** Rating scale did not contain items that permitted faculty to indicate registrant's level of competence in performing clinical procedures.

Table 8

Complex Correlations Between Rating Factors
and Components of Examination Performance

Examination Component	Rating Factors	Multiple Correlation *
Identification and Classification of Physical Findings	Data Gathering, Clinical Judgment	> .29
Patient Counseling and Instruction	Interpersonal Skills, Clinical Judgment	> .33
PMP Data Gathering	Data Gathering, Clinical Judgment	> .37

* All correlation coefficients are significant beyond the .05 level

While the pattern of significant correlation coefficients between examination performance and clinical competence ratings provided evidence for the convergent and discriminant validity of the Certifying Examination, the magnitude of these correlation coefficients was modest. An analysis of the rating scale itself suggested that several modifications could be made in the format which might enhance the sensitivity of this measuring instrument. (One of the difficulties that has long plagued rating scales has been their inability to distinguish different levels of competence within the same individual along different rating dimensions.) For this reason, significant modifications were made in the rating scale and a replication of this study carried out in relation to the 1974 Certifying Examination.

The following modifications were made in the rating scale used in the replication.

1. the number of competency statements to be rated was increased from 40 to 52;
2. raters were asked to indicate the degree of confidence they had in the rating given to each physician's assistant for each competency statement;
3. raters were offered the option of indicating that for a given competency statement they did not feel able to judge the physician's assistant.

The results of analyses between clinical competence ratings and examination scores did not reveal a marked increase in the convergent or discriminant validity of the Certifying Examination. In addition, the factor analysis of rating scale items did not yield markedly different factors from those that had been observed in the first study. However, it was found that confidence in the ratings provided differed significantly among the three groups of individuals (program administrators, program faculty, and employers) who provided ratings on physician's assistants. In general, program administrators were significantly less confident in their ratings of clinical competence than were program faculty or employers. This finding suggests that certain types of individuals should not be asked to provide ratings of competence where they have limited observations and/or experience on which to base their ratings.

(A detailed report of the 1973 study was published in the Journal of Medical Education, March 1976.)

EXPANDED ELIGIBILITY - 1974 CERTIFYING EXAMINATION

In addition to graduation from an approved educational program, eligibility for the 1974 Certifying Examination was expanded to include individuals qualifying on the basis of work experience. Individuals wishing to so qualify were required to have a high school diploma or an equivalency certificate and four years of medical clinical experience in primary care as a physician's assistant or nurse practitioner since January 1, 1970. The work experience outlined on each application form was verified by contacting the physicians whose names and addresses were provided. Each physician was required to provide a detailed description of the health care functions performed by the applicant as well as an estimate of the frequency with which each was performed. These details of an applicant's employment history were then evaluated in relation to specific criteria that had been established in advance by the Eligibility Committee. In instances where an applicant was not clearly eligible or ineligible, the application form and supporting documents were reviewed individually by members of the Eligibility Committee.

Announcements of the eligibility requirements were placed in the major primary care medical journals and newsletters. As a result, 550 completed application forms were received from individuals wishing to qualify on the basis of work experience. Of this number, 150 met the established eligibility requirements and 116 actually sat for the Certifying Examination. The most frequently encountered factors in failing to meet eligibility requirements involved less than the required four years of clinical experience and health care functions that did not include making patient management and disposition decisions.

1974 PHYSICIAN'S ASSISTANT CERTIFYING EXAMINATION

The written component of the 1974 Certifying Examination followed the same format as was used in 1973: that is, a multiple-choice examination designed to assess the candidate's knowledge and skill in applying knowledge related to clinical material presented in printed and pictorial form, and patient management problems involving simulated clinical cases in adult and pediatric medicine designed to assess the candidate's skill in gathering pertinent information about patients and in making appropriate management decisions. The number of items included on both portions of the written examination was increased on the 1974 Examination. In addition, multiple-choice questions using pictorial material involved the identification and interpretation of physical findings instead of the identification and classification of findings as had been the case in the previous year.

In addition to the written component of the examination, each candidate underwent assessment of his/her physical examination skills using the standardized behavioral checklists that had been developed as part of a research and development project. This assessment consisted of an evaluation of the candidate's proficiency in performing five components of the physical examination (i.e., heart, lungs, eyes, abdomen, and neurologic). Test centers were established on the two days following the written examination and candidates were given individual appointments for assessment. The candidate performed three examinations on one patient while being observed by a physician examiner, and performed the remaining two examinations on another patient while being observed by another physician examiner. The total assessment time for each candidate was approximately forty minutes. At the completion of each examination, the candidate was asked to describe any abnormalities found. However, due to the variability in abnormal findings among patients and the limited sample of abnormal findings that was available, this portion of the assessment procedure was not used for scoring purposes. (A detailed report concerning the development and field testing of these checklists is provided in the Project Report, 1973-74.)

Physician examiners were appointed by program directors at the various test centers Prior to the examination, three regional orientation meetings were held with examiners in order to acquaint them with the assessment objectives, procedures, and guidelines for patient selection. In addition, proctors were appointed for these special test centers to ensure the adequacy of the logistical arrangements and to maintain the security of the assessment procedure itself. Since written reports from the chief proctors and physician examiners did not disclose any problems that would jeopardize the integrity of this assessment procedure, the physical examination assessment component of the Certifying Examination was used for actual scoring purposes Each of the five physical examination components was weighted equally, and performance on this component of the examination accounted for twenty-five per cent of a candidate's total score.

STATISTICAL ANALYSIS - 1974 CERTIFYING EXAMINATION

The 1974 Certifying Examination was administered on December 11-13 to 1303 candidates in 49 test centers. The statistical properties of the 1974 examination program closely paralleled those encountered in 1973. Although the number of candidates still enrolled in educational programs increased to twenty-three per cent in 1974 as compared to eleven per cent in 1973, the typical examinee continued to be an individual who had already completed a formal educational program and acquired up to two years of postgraduate clinical experience. Seventy-five per cent of the formally trained examinees received their training in programs from 13 to 24 months in length, and ninety-two per cent had been involved in health care delivery prior to being trained as a physician's assistant or nurse practitioner. Of this number, eighty-four per cent had been involved in patient contact health care delivery. Of those examinees with prior experience in health care delivery, fifty-eight per cent had more than four years of experience. These data are summarized in Table 9 on the following page.

Table 9

Physician's Assistant Certifying Examination

Description of Examinee Populations

	1973	1974
I. Percentages of total respondents:		
A. Current status		
Currently in educational program	11%	23%
Currently employed or graduated	89	77
B. Type of training		
Physician's assistant/associate	62	71
Medex	29	15
Nurse	9	5
Informally trained	--	9
C. Amount of experience as a P.A.		
None	19	25
0-2 years	72	62
More than 2 years	9	13
II. Percentage of respondents with formal training (informally trained candidates not included)		
A. Length of educational program		
4-12 months	22	13
13-24 months	65	75
25-36 months	10	7
More than 36 months	3	5
B. Experience in health care delivery prior to educational program		
Yes	91	92
No	9	8
1. Type of experience		
Technical	11	12
Patient contact	86	84
Other	3	4
2. Length of experience		
0-1 year	4	4
1-2 years	10	11
2-4 years	30	27
More than 4 years	56	58

Statistical analysis of the 1974 examination indicated that the average difficulty level and reliabilities for the various components of the examination were comparable to statistics derived from the 1973 examination. The one exception is the average difficulty level of the multiple-choice questions which indicated that those developed for the 1974 examination were somewhat more difficult than the MCQ's developed for the 1973 examination. (In 1973, the average difficulty level equaled .64; in 1974, this figure was .57.) The composite reliability for all examination components increased from .89 in 1973 to .93 in 1974. This modest increase in reliability is attributable to the larger number of items contained on the 1974 examination. Comparative statistics for the first two Certifying Examinations are provided in Table 10.

An analysis of the inter-correlations among the various examination components yield correlation coefficients ranging from .13 to .53. The correlations among the portions of the written examination ranged from .37 to .53, and closely resembled those encountered on the 1973 examination program. These modest correlations suggest that the various portions of the written examination are measuring different aspects of competence. As might be anticipated, the correlation between the written portions and the physical examination assessment portion of the Certifying Examination were lower, ranging from .13 to .28. (These inter-correlations are summarized in Table 11.) Given the magnitude of these inter-correlations, it is evident that one cannot predict a candidate's physical examination skills by knowing his score on the written portion of the examination. Thus, it appears reasonable to conclude that the performance component of the examination was contributing significant information concerning the competence of candidates that could not be inferred from performance on the written portions of the examination. Moreover, the wide variation in scores on the physical examination assessment component indicated a wide range of proficiency among the 1303 candidates who sat for the 1974 examination. This range of scores was greater than that observed for the written component of the examination.

Table 10

Comparative Examination Statistics (1973-1974)

Exam Component	Number of Items		Average Difficulty		Reliability		Composite Reliability	
	1973	1974	1973	1974	1973	1974	1973	1974
MCQ	147	216	.64	.57	.83	.89	.89	.93
PMP - Data Gathering	422	554	.79	.75	.80	.89		
PMP - Management/ Therapy	37	284*	.82	.78	.75	.76		
Physical Examination Skill Assessment	--	212	--	.67	--	.91		

* The increase in the number of items reflects the fact that the selection of laboratory studies was included in the management sub-score rather than in the data gathering score as had been the case in 1973.

Table 11

INTER-CORRELATIONS AMONG SUB-SCORES
1974 CERTIFYING EXAMINATION

	PMP (Mgmt./Therapy)	PMP (Data/Gath.)	Physical Exam Assessment
MCQ	.53	.41	.28
PMP (Mgmt./Therapy)		.37	.12
PMP (Data/Gath.)			.13

In addition to the above analyses, a comparison was made of the performance of informally and formally trained candidates on all components of the Certifying Examination. In all instances, candidates who had not graduated from formal educational programs scored significantly lower than did candidates who had been trained in formal educational programs. With respect to examinees who had been trained in formal educational programs, candidates from physician's assistant and nurse practitioner programs scored equally well on the written portions of the examination. However, candidates from Medex programs scored significantly lower on the written examination when compared with the two other groups of formally trained examinees.

With respect to the physical examination skill assessment, candidates from physician's assistant programs scored significantly higher than candidates from Medex or nurse practitioner programs. There was no significant difference in the performance of Medex and nurse practitioner candidates on the physical examination skill assessment. In interpreting the findings concerning nurse practitioners, it should be remembered that the number of nurse practitioners sitting for this Certifying Examination was not only very small (N=69), but was probably not representative of nurse practitioners in general. The number of nurse practitioner candidates is probably somewhat less than 10 per cent of the total number of nurse practitioners trained in any given year.

As the 1974 Certifying Examination incorporated a new evaluation methodology for the assessment of physical examination skills, it was hoped that research on audiovisual simulation would result in the development of yet additional new evaluation methodologies for the assessment of another important area of physician's assistant competence - interpersonal skills.

Setting Examination Standards

The method that has become most widely used for setting pass-fail levels on standardized examinations involves the use of norm-referenced procedures. When these procedures are employed, a pass-fail level is determined on the basis of how well examinees perform in relation to one another. Within the last few years, a number of articles have appeared in the literature describing different procedures for setting examination standards. These procedures have been described as criterion-referenced and differ from norm-referenced procedures in that the pass-fail level is established as a function of a predetermined standard rather than as a function of the relative performance of examinees in relation to one another.

In order to investigate the nature and stability of these criterion-referenced procedures, two were selected and used experimentally by test committees for the 1973 Physician's Assistant Certifying Examination. The purposes of this study were to determine:

1. whether the pass-fail levels for comparable samples of test content would vary depending upon the standard setting procedure used to set this level;

2. whether for each of two standard setting procedures, the pass-fail level for sample test content would vary depending upon the group of judges whose decisions were used to set the standard;

3. whether the average of judgments made by individuals within each group would differ significantly from the group judgments arrived at by consensus.

The two standard setting procedures selected were chosen because both were designed to set an examination standard in relation to a "minimum acceptable level of proficiency," and were amenable to use with individual or group judgments. Although the two procedures differed in terms of the specific techniques used to develop the pass-fail level, both procedures required groups of judges to review each test item and make a decision concerning how well a minimally qualified candidate should perform in relation to it. Judgments concerning individual test items were then combined in such a way as to provide an overall pass-fail level.

Two groups of judges were selected at random from among the various test committees. Each group utilized both standard setting procedures. On the basis of the analyses performed, the following conclusions were reached:

1. Different groups of judges using the same criterion-referenced procedure in relation to the same examination content did set similar overall pass-fail levels.

2. When the magnitude of agreement between groups of judges using the same criterion-referenced procedure was estimated on an item-by-item basis, product-moment correlation coefficients of .40 - .45 were obtained. This indicates a modest positive correlation between groups of judges with respect to the judgments they made about individual test items when using the same criterion-referenced procedure.

3. The averages of judgments concerning examination standards made by individuals within each group did not differ significantly from the consensus judgments of the group as a whole.

4. The two criterion-referenced procedures used in this study did, however, yield significantly different pass-fail levels for comparable samples of test content.

This study was replicated using the 1974 Certifying Examination and different groups of judges. The purpose of this replication was to validate the conclusions reached in the first study The preliminary statistical analysis indicates that the two criterion-referenced procedures continue to result in different pass-fail levels In contrast to the previous study, however, it appears that the two groups of judges set different pass-fail levels even when using the same standard setting procedure.

Although it is premature to formulate any definitive conclusions regarding the nature of criterion-referenced procedures for setting examination standards, certainly the results of the second study raise a question concerning the stability of either procedure in setting a consistent pass-fail level for a given sample of examination content. While these findings probably do not differ from those that would be observed with use of norm-referenced standard setting procedures, they do challenge the assumption that has been made by many that criterion-referenced procedures constitute a more objective and stable means of setting examination standards.

It is perhaps reasonable to think that the pass-fail level set by whatever standard setting procedure will vary significantly as a function of the individuals who are selected to set that standard. For this reason, it is conceivable that decisions to use one standard setting procedure as opposed to another will be made largely on philosophical grounds rather than upon statistical data to support one approach over another. A further problem which must be faced is that criterion-referenced procedures are applicable only to multiple-choice and matching type test items, and cannot be used with items on patient management problems which constitute more than half of the Physician's Assistant Certifying Examination. (A report of this research study was published in Educational and Psychological Measurement, Spring 1976.)

The following sections of this report contain a detailed description of activities carried out during Phase II of this project including: (a) a study of audiovisual simulation techniques for assessing interpersonal skills; (b) a validation study of the 1974 Certifying Examination involving medical record audits; and (c) an analysis of the 1975 Certifying Examination.

AUDIOVISUAL SIMULATION STUDY

Introduction

When historians and social scientists attempt to synthesize the accomplishments of the first three quarters of the twentieth century and characterize the evolution of civilization as manifest during that period, it will no doubt be man's scientific and technologic advances which will occupy the major part of their narrative. As literature, philosophy, and the arts have constituted the humanistic legacy of man's achievements in the past, this century will no doubt be viewed as a watershed in man's scientific heritage. To the extent that science has preoccupied our creative and pragmatic energies in this century, the humanistic dimensions and their potential contribution to our quality of life have been overshadowed.

In the health professions, this has resulted in a heightened capability to comprehend and master the complex biophysical processes which determine health status. In large measure, however, health professionals have lost sight of and insight into the psychosocial components of man and his environment and the implications of these components for health. Armed with the powerful weapons of science, health professionals have neglected the humanistic aspects of problem-solving and have not fully developed the skills necessary for effective human interface.

Within the last fifteen years, studies whose focus has been the definition of professional competence have begun to highlight a number of dimensions which have renewed interest in what might be considered the humanistic components of competence. As an example, in 1960, the American Institutes for Research conducted a critical incidents study of intern and resident performance for the

National Board of Medical Examiners. The purpose of this study was to identify critical incidents of physician performance which were felt to have a positive or negative impact on the quality of health care. Interwoven among the over 3,000 critical incidents which were gathered from this study are examples of behavior which might be described as interpersonal and communication skills. In other studies of clinical competence using somewhat different methodologies, medical specialty boards and societies have also included these dimensions as important components of physician competence. It is perhaps not surprising, therefore, that definitions of clinical competence for the primary care physician's assistant have also included interpersonal and communication skills.

Since these skills were felt to be important competency components for the primary care physician's assistant, a project was undertaken to develop a detailed definition of these skills and to design and validate several evaluation methodologies which could assess these components of competence for purposes of certification in the profession.

Criterion Development and Measurement Issues

It was felt that the assessment of interpersonal and communication skills should be undertaken within the context of patient interviewing and counseling activities. The first step in this process was to identify performance criteria that could serve as the basis for evaluating the behavior of physician's assistants. In order to identify appropriate performance criteria, a project committee was appointed. The committee was composed of individuals who were active in teaching and evaluating interpersonal and communication skills within physician's assistant training programs. Membership was inter-disciplinary and included representaion from medicine, surgery, pediatrics, psychiatry, clinical psychology, and nursing.

Performance criteria associated with patient interviewing and counseling were specified as presented in Table 1 below.

Table 1

Performance Criteria for Patient Interviewing and Counseling

I. Content of Questions or Statements

A. Uses questions or statements that will provide pertinent data about the patient.

B. Does not use questions or statements which will frighten, anger, insult, or embarrass the patient.

C. Questions patient about inconsistencies.

D. Uses language appropriate to the educational level and background of the patient.

E. Avoids use of medical "jargon."

F. Collects sufficient data before terminating the interview.

II. Format of Questions

A. Selects question formats which will yield the greatest amount of data with the greatest accuracy. (These will generally be open-ended or reflective questions.)

III. Cue Sensitivity

A. Follows up pertinent verbal and non-verbal cues provided by patient.

B. Encourages patient to express pertinent information and feelings.

IV. Patient Rapport

A. Appropriately answers patient questions (giving appropriate information without frightening, angering, or embarrassing the patient).

continued.....

B. Provides patient with appropriate information (in terms of accuracy and amount).

C. Deals appropriately with different patient personalities.

D. Gives reassurance and support where appropriate.

E. Appropriately expresses understanding and/or empathy.

F. Maintains a non-judgmental, respectful attitude toward the patient.

G. Maintains eye contact with the patient.

H. Acknowledges patient comments and concerns.

I. Uses gestures, facial expressions and posture which convey attention and interest.

J. Does not interrupt the patient.

K. Terminates interview in an appropriate manner.

(1) Asks if there's any additional information patient wants to provide.

(2) States immediate disposition of what is to happen next.

In reviewing the performance criteria specified in the above and considering the nature of the activites to which these criteria were related, it was felt that three dimensions were crucial to these activities and should be captured to the extent possible in the design of evaluation methodologies. Since the course of a patient interviewing or counseling session is determined in large measure by the nature of the verbal and non-verbal cues presented by the patient, it was felt that audiovisual stimuli would be an important component of any evaluation methodology. In addition, since it is probable that considerable variation will exist among health professionals in terms of their approach to a specific interviewing or counseling session, it was felt that capacity for branching logic and multiple pathways would be another dimension that should

be included in an evaluation approach. Finally, the need to provide audiovisual stimuli and branching logic led to the identification of interactive capability as a third relevant dimension in the evaluation process.

Associated with any evaluation enterprise is the need to identify not only relevant performance criteria and evaluation dimensions but also the need to identify measurement problems and aspects of economic and logistic feasibility that will influence the selection and design of appropriate evaluation methodologies.

In relation to measurement problems, two seemed particularly relevant: 1) standardization of the content of an evaluation technique and 2) development of scoring strategy that numerically reflects the quality of individual performance criteria.

To the extent that observation of live performance is used as one evaluation approach, the problem of standardization is a particularly difficult one. Moreover, the observation of live performance results in generally higher costs and more complex logistics associated with its administration. Unfortunately, our experience to date with observation of live performance suggests that while the measurement and logistic problems can be overcome, the cost of doing so raises major questions about the ultimate feasibility of utilizing observation of live performance as an ongoing mechanism for evaluating professional competence. However, if these problems can be brought within acceptable levels of tolerance, it remains that the evaluation of actual performance yields more accurate assessment of performance than any indirect assessment technique be it paper and pencil, audiovisual simulation, or computer simulation.

Thus, the challenge that faces agencies with responsibility for evaluating professional competence is the development of indirect evaluation methodologies that are capable of predicting with reasonable accuracy the performance of health professionals as it would be observed in an actual clinical setting. In this context, the measurement of live performance can be used as a criterion measure against which to assess the validity of indirect evaluation methodologies. To the extent that one or more of these methodologies yields significant positive correlations with actual performance, their validity can be judged.

Evaluation Methodologies

It was concluded that the basic research design for this study would utilize the observation of live performance in patient interviewing and counseling sessions as the criterion measure against which to assess the validity of several evaluation approaches involving three types of simulations.

Several studies in the literature have reported the use of interaction observation forms for evaluating patient interviewing and counseling skills. (1-6) These observation forms have contained specified categories of behavior which were to be observed and recorded by a trained observer as a physician-patient interaction was under way. Since studies involving a limited number of behavioral criteria had demonstrated that acceptable reliability coefficients could be obtained using this evaluation approach, it was felt that a similar approach should be developed and reliability studies conducted in order to develop the criterion measures of performance for this study.

However, two aspects of this study differed significantly from previous efforts. First, the number of behavioral criteria that needed to be included on the observation form was greater than the number of criteria that had been utilized in previous studies. Thus, a significant question was whether trained

observers could learn a large number of behavioral categories, quickly classify observed behavior into one or more categories, and record the decisions accurately on the observation form without jeopardizing the accuracy and reliability of their observations. Second, although previous studies had focused on the description of behavior during interviewing and counseling sessions, none of these studies had attempted to develop a scoring strategy which could describe the quality or appropriateness of the behavior that had been observed. Thus, the problem of placing a value judgment on observed behavior had to be faced in the conceptualization of a logical and reasonable scoring strategy.

Three indirect evaluation approaches were included in this study. Each approach differed from the others in the extent to which relevant dimensions (i.e., audiovisual stimuli, branching logic, and interactive capability) could be incorporated. The three evaluation approaches included:

a. problem-centered multiple-choice questions

b. branching, paper-pencil simulations

c. interactive audiovisual simulations.

All evaluation approaches involved the presentation of problem situations in which the subject was required to make decisions concerning the question or comment to the patient which would be most effective in dealing with the specific problem situation. However, the multiple-choice questions did not provide for audiovisual stimuli, interaction, or branching capabilities. A sample of the problem-centered multiple-choice questions is provided in Table 2.

Table 2

Sample Multiple-Choice Questions

CASE B (Questions 5-11):

Kathy D., a 14-year-old eighth grader, comes to the office because of two missed menstrual periods. During the history you learn that Kathy's menarche occurred at age 12 and for the past year menstruation has been regular every 28 days with the exception of the past two months. You further learn that Kathy has been dieting rigidly over the past 4 months in an effort to lose 15 pounds.

From your viewpoint she now looks too thin, her clothes hang, her cheekbones show, her hair and eyes lack luster, and she is less vivacious than when you saw her a year ago. She is 5'4" and weighed 110 pounds before her diet.

She tells you that she has lost 12 pounds and only has three to go. She says, "This diet has been the hardest thing I've ever done, but look at me...Don't I look terrific?"

5. The best immediate response is

(A) "It must have been hard to lose all that weight in just a few months."
(B) "You really look too thin, Kathy."
(C) "...terrific?"
(D) "What made you go on the diet in the first place?"

6. When you ask Kathy how she feels in general, she tells you that lately she has been tired more than usual. She also says that she notices the cold more than in the past, and finds herself wearing a sweater at times that she would normally be comfortable with just a dress or pants and a shirt. She says, "This being tired all the time is such a drag. I get more than enough sleep, but I'm still tired. How come?" The best response would be

(A) "I expect one reason is that you've cut down on a good source of energy - food "
(B) "It may be that your sleep is not restful."
(C) "It's not unusual for adolescents to feel tired much of the time. Your body is actively maturing, which is hard work."
(D) "Before drawing any conclusions, I'd like to examine you and do some tests."

7. When you ask Kathy how her parents feel about her dieting, she tells you that it's given them one more thing to harp on. She says, "You know my parents have been all over me since I entered junior high. If it's not my grades, it's my language, my clothes, my messy bedroom, my laziness, and now it's my eating or lack of it. I wish they'd bug off." The best response is

(A) "I imagine your parents are concerned about your welfare."
(B) "It must be difficult to live in such a negative atmosphere."
(C) "Why don't you bring your parents in with you for a conference?"
(D) "It sounds as though your life changed after you entered junior high."

8. Kathy continues talking about her parents and says, "Actually, I'm here because my mother made me come. She got all upset when she found out I missed a couple of periods. She made such an issue out of it. You'd think I was pregnant. What's the big deal about a couple of missed periods?" The best response is

(A) "Did you mother have reason to think you might be pregnant?"
(B) "What does a missed period mean to you?"
(C) "Your mother is probably worried that your dieting resulted in the missed periods."
(D) "Sometimes missed periods can be a warning signal of problems other than pregnancy."

The branching paper-pencil simulations were designed so that each simulation began with a case description of the patient and offered a variety of questions and statements from which the subject could select the one he felt was most appropriate for that point in the simulated interaction. As each question or comment was selected, the patient's response could be read by using a latent image pen to develop the invisible print. At the end of each patient's response, instructions were provided for turning to the next series of options and selecting the next question or statement. Since each simulation involved branching logic, subjects were able to follow different pathways of interaction with each simulated patient. Moreover, the patient's response to each question or statement provided the basis from which a thread of continuity could be established in the interaction. Thus, this evaluation approach was able to capture two of the three relevant dimensions of the interviewing and counseling process: branching logic and interactive capability. Appendix A contains a sample paper-pencil simulation with all patient responses exposed for illustrative purposes.

The use of interactive audiovisual simulations necessitated the conceptualization and development of an innovative system which would be capable of providing random access to audio and visual stimuli as well as interaction with the simulated patient by means of a branching logic and videotape program. A prototype Interactive Audiovisual Simulation System had been developed in connection with an earlier project and was incorporated into the design of this study.

The unique aspects of this Interactive Audiovisual Simulation System involve the interface between random access slide projection and random access videotape display, flexible branching logic, rapid and accurate search and retrieval, and a hard copy printout of selections made.

In using this system, the subject sits in front of a display console and interacts with a patient who has been filmed on videotape according to a branching logic script. The subject advances through each nodal point of the branching simulation by means of selecting from a series of options the question or statement he believes appropriate for that moment in the interaction with the patient. The code number of the option is entered into the system which then retrieves and plays the videotape segment containing the patient's response to the subject's selection. Based upon the audiovisual cues presented, the subject can then decide whether to pursue the same line of questioning or to begin a new aspect of inquiry. Some pathways lead to the successful resolution of the interviewing or counseling session, while others result in a premature termination of the session because an ineffective or inappropriate approach to the patient has been selected or because the interviewer chooses to terminate the interaction at an earlier point.

In order to develop interaction observation forms for assessing performance in a live interview and counseling session, the performance criteria that had been specified by the project committee were reviewed and categories of appropriate and inappropriate behavior consistent with each performance criterion were identified. As a result, 28 categories of behavior were developed. These categories are listed in Table 3.

Table 3

Behavioral Categories

Interaction Observation Form

Question Content/Format

1. Asks non-case pertinent, open-ended question
2. Asks case pertinent, open-ended question
3. Asks non-case pertinent, close-ended question
4. Asks case pertinent, close-ended question
5. Asks case pertinent, inappropriately leading question
6. Reflects patient comment or question as a probe
7. Asks repetitious question (i.e., the same question asked multiple times)
8. Uses continuous questioning (i.e., asks more than one question at the same time)

Reactions to Patient Behaviors

9. Inappropriately interrupts patient/abruptly shifts focus of interaction
10. Encourages patient to continue talking
11. Uses medical jargon
12. Follows up pertinent verbal or non-verbal cues (new information)
13. Fails to follow-up pertinent verbal or non-verbal cue (new information)
14. Repeats own prior question/comment verbatim
15. Clarifies own prior question/comment
16. Appropriately summarizes/synthesizes patient comments
17. Inappropriately summarizes/synthesizes patient comments

continued.....

Provision of Support/Information

18. Expresses understanding of what patient has said
19. Responds to patient comments with inappropriate affect (e.g., disinterest, judgment, disapproval)
20. Reassures patient appropriately
21. Reassures patient inappropriately
22. Answers patient's questions directly
23. Answers patient's questions evasively
24. Gives advice/instructions appropriately
25. Gives advice/instructions inappropriately
26. Provides appropriate information
27. Provides inappropriate information
28. Expresses empathy

The above behavioral categories were then formulated into an observation form as illustrated in Appendix B. The observation form was designed so that a trained observer could record the behaviors which were observed in the sequence in which they took place. In instances where the term "appropriate" appears in relation to a behavioral category, it was felt that the appropriateness or inappropriateness of the behavior would be a function of the specific patient and his or her problems, personality characteristics, and purpose of the interview or counseling session.

Development of Patient Protocols

In order to develop the clinical cases which would serve as the content of the various evaluation approaches, the project committee was asked to generate a total of 14 patient protocols. In their selection of case materials,

they were asked to focus upon patients whom it might reasonably be expected would be seen by primary care physician's assistants. For this purpose, committee members were asked to rely upon case materials drawn from actual patient files. In order to guide their development of these patient protocols, each project committee member was given a patient protocol outline which described the categories of descriptive data they should develop for each clinical case. Each resulting patient protocol consisted of a sufficiently comprehensive yet succinct description of the patient's current problems and life situation, affective profile, medical and psychosocial histories, clinical setting, and behaviors that physician's assistants should and should not engage in. A sample of the patient protocol outline is supplied in Appendix C.

All patient protocols were reviewed by the entire committee and modified as necessary in order to develop a final protocol which would be a realistic yet challenging clinical case for the physician's assistant.

Simulation Development

The authors of each patient protocol were then asked to write up to seven problem-centered multiple-choice questions per protocol. These test questions were reviewed by the project staff and subsequently by the entire committee in order to maximize the face validity of each multiple-choice question. A total of 82 such questions were finally selected for inclusion in the series of multiple-choice questions used for this study.

The project committee was then asked to select protocols which they felt were rich in content and opportunities for the evaluation of skills in patient interviewing and counseling. Seven such protocols were identified, and each member of the committee was assigned the responsibility of developing a branching paper-pencil simulation script based upon one protocol. The resulting

seven scripts were reviewed in detail and modified by the committee at a series of subsequent meetings. In developing the options to be presented at each decision point in the script, an effort was made to present questions and comments which reflected different degrees of appropriateness. In addition, the content of all options was designed so as to permit as wide a variation in interviewing approach as possible within the constraints of the paper-pencil methodology. Of the seven scripts reviewed by the committee, five were selected for further development and use with the paper-pencil simulations and interactive audiovisual simulations. The following is a selected sample of the clinical case materials contained in the various patient protocols:

a. unwed pregnant teenage girl facing conflict with her family regarding what she should do about her pregnancy as well as her own uncertainty about the course of action she should take;

b. a male patient who is recovering from a myocardial infarction and having difficulties coping with his altered self-image and the need to modify his life style;

c. a married woman with a prolonged upper respiratory infection, secondary to marital problems;

d. an obese male who is on a weight control program and who is having difficulty coping with a new self-image and the demands placed upon him by his altered peer relationships:

e. a woman who is denying the fact that her first child has been born with Down's Syndrome and who has come to the practice seeking "another medical opinion";

f. a pregnant woman who is concerned that her mild hypertension in the last trimester of pregnancy may result in an unsuccessful outcome as has been the case in her first two pregnancies;

g. an elderly man who has a series of physical and psychological problems secondary to the effects of the aging process.

Scoring Strategy

The twenty-eight categories of behavior developed for the interaction observation form were used not only to classify behavior observed during the live interviews but also to classify options selected on the multiple-choice questions, paper-pencil simulations, and interactive audiovisual simulations. This procedure was followed so that comparisons could be made among the evaluation methodologies.

These categories of behavior were reviewed and related behaviors were clustered together. This clustering resulted in the identification of the five scoring variables.

a. efficiency

b. non-directiveness

c. facilitation of expression/rapport - support

d. enhancing patient understanding

e. responsiveness to patient behaviors

Each scoring variable consisted of a positive and a negative dimension. Behaviors within each scoring variable were assigned to one or the other dimension as a function of whether or not it was felt to contribute to the communication goal used to describe the particular scoring variable. Table 4 illustrates the positive and negative exemplars of behavior associated with each scoring variable.

Table 4

Scoring Variables and Behaviors
Included in Each

Variable 1: EFFICIENCY

Positive	Negative
Asks case pertinent, open-ended question	Asks non-case pertinent, open-ended question
Asks case pertinent, close-ended question	Asks non-case pertinent, close-ended question
Follows up pertinent verbal or non-verbal cues	Asks repetitious question
	Fails to follow-up pertinent verbal or non-verbal cues

Variable 2: NON-DIRECTIVENESS

Positive	Negative
Asks case pertinent, open-ended question	Asks case pertinent, close-ended question
Reflects patient commment or question as a probe	Asks case pertinent inappropriately leading question
Encourages patient to continue talking	Asks continuous questions
	Inappropriately interrupts patient/abruptly shifts focus of interview
	Repeats own question verbatim

Variable 3: FACILITATION OF EXPRESSION/ RAPPORT - SUPPORT

Positive	Negative
Reflects patient comment or question as a probe	Inappropriately interrupts patient/ abruptly shifts focus of interview
Encourages patient to continue talking	Uses medical jargon
Appropriately summarizes/synthesizes patient comments	Inappropriately summarizes/synthesizes patient comments
Expresses understanding	Responds with inappropriate affect
Reassures patient appropriately	Reassures patient inappropriately
Expresses empathy	

Variable 4: ENHANCING PATIENT UNDERSTANDING

Positive	Negative
Clarifies own prior question/comment	Uses medical jargon
Gives advice/instructions appropriately	Repeats own question verbatim
Answers patient questions directly	Gives advice/instructions inappropriately
Provides appropriate information	Answers patient questions evasively
	Provides inappropriate information

Variable 5: RESPONSIVENESS TO PATIENT BEHAVIORS

Positive	Negative
Reflects patient comment or question as a probe	Inappropriately interrupts patient/abruptly shifts focus of interview
Encourages patient to continue talking	Fails to follow-up pertinent verbal or non-verbal cues
Follows up pertinent verbal or non-verbal cues	Repeats own question verbatim
Expresses understanding	Responds with inappropriate affect
Reassures patient appropriately	Reassures patient inappropriately
Answers patient questions directly	Answers patient questions evasively
Expresses empathy	Provides inappropriate information

The specific behaviors included in each of the scoring variables was determined by the combination which yielded the highest intra-class correlation coefficient for a given variable. Thus, before determining final composition of each scoring variable, a number of intra-class correlation coefficients was calculated, eliminating behavioral categories that resulted in considerable variation in observations among the raters.

After the number of scoring variables and their composition had been determined, attention turned to the development of a scoring formula which could provide numerical values reflective of the quality of the behaviors associated with each scoring variable. The scoring formula was based upon a number of principles which were formulated in advance. For example, it was felt that the score for each variable should reflect the relationship between the number of positive and the number of negative behaviors observed. Moreover, it was felt that the number of negative behaviors should be weighted more heavily than the positive behaviors since their impact upon the interactive process is probably greater per instance of behavior than is the impact of each positive behavior. With these principles in mind, a number of hypothetical data sets were scored using a preliminary scoring formula. This preliminary formula was modified somewhat to better reflect the principles stated above, and the following represents the scoring formula finally selected:

$$X = \frac{\Sigma \text{pos.} + 1}{1.5\ (\Sigma \text{neg.}) + 2}$$

where,

Σpos. = sum of the positive behaviors

Σneg. = sum of the negative behaviors

Utilizing the above formula, the following scores would be obtained from the data set presented below.

Σ pos.	Σ neg.	X
0	10	.059
0	5	.105
1	10	.118
0	1	.286
1	1	.571
5	5	.631
10	10	.647
1	0	1.0
5	0	3.0
10	1	3.14
10	0	5.5

All items contained in the multiple-choice questions, paper-pencil simulations, and audiovisual simulations were reviewed by a member of the project staff and classified into one or more behavioral categories. Thus, the profile of behaviors produced by each subject could be compared across all evaluation methodologies. The scoring formula presented above was used to calculate scores for each variable on all evaluation methodologies.

Reliability Studies - Interaction Observation Form

Before using behavior in a live interactive setting as the criterion measure in this study, two inter-rater reliability studies of the interaction observation form were conducted. It was felt that the best indicator of the stability of the measurements derived from this form would be the intra-class

correlation coefficient. This statistic is an estimate of the average agreement among multiple observers providing data on the same interviewing or counseling session.

In order to achieve a maximum standardization of the content of these live interviews in the first study, four patient protocols were selected and individuals were recruited to portray each role. The individuals selected were non-professional actors from community acting groups and the University of Pennsylvania School of the Performing Arts. Each actor and actress was given a period of orientation and training in portrayal of their respective roles.

Eight physician's assistants from a local training program volunteered to participate in the reliability study by interviewing the programmed patients. Each physician's assistant interviewed two of the four programmed patients, and each interview was videotaped using two hidden cameras. A total of sixteen videotaped interviews was available for analysis.

The project committee was brought to Philadelphia where they were given a half-day training program in the use of the interaction observation form. They were also asked to critique the forms and offer suggestions for modification. These suggestions were incorporated into the forms, and the committee was then asked to view four videotaped interviews and to individually classify the behaviors on the modified interaction observation form. Committee members viewed each videotape simultaneously but individually completed the form. Thus for each of the videotaped interviews there were six interaction observation forms.

The twenty-eight behavioral categories on the observation form had been clustered into five scoring variables for purposes of evaluating physician's assistant behavior. These scoring variables were labeled as follows:

a. efficiency

b. non-directiveness

c. facilitation of expression/rapport - support

d. enhancing patient understanding

e. responsiveness to patient behavior.

An intra-class correlation coefficient was calculated for each of the five variables. For this first reliability study, these correlation coefficients ranged from .70 to .87. It was felt, however, that with further training of the raters these reliability coefficients could be increased and provision was made to conduct a second reliability study.

The second reliability study followed the same format as the first, but a more systematic approach was used in training the committee members to use the behavioral categories. A different set of four videotaped interviews was shown.

An intra-class correlation coefficient was calculated for each scoring variable and, as had been hoped, four of the five reliability estimates increased. In the one instance where there was a slight decrease in the reliability estimate, it was not felt to be a significant difference. The reliability estimates for the first and second reliability studies are presented in Table 5. In the first study, the intra-class correlation coefficient was calculated using six observers, while in the second study five observers were present.

The results of the second reliability study suggested that the interaction observation form did yield sufficiently reliable estimates of performance to warrant its use as a criterion measure in the overall research design.

Table 5

Reliability Estimates[a] of
Interaction Observation Form

Variable	Study #1[b]	Study #2[c]
Efficiency	.72	.86
Non-Directiveness	.70	.93
Facilitation of Expression/ Rapport - Support	.77	.82
Enhancing Patient Understanding	.87	.82
Responsiveness to Patient Behaviors	.70	.84

[a]based upon intra-class correlation coefficient

[b]using 6 judges

[c]using 5 judges

Research Design and Data Collection

Subjects for this study were drawn from three physician's assistant training programs. All subjects were volunteers who were paid a modest honorarium for their participation in the study. A total of 59 subjects participated, nineteen at site one and twenty each from sites two and three. All subjects were evaluated using all evaluation approaches.

In order to control for the sequence in which subjects would experience the various methodologies, the subjects at each site were randomly assigned to one of four groups. The sequence in which each group experienced the evaluation methodologies is presented below:

Group A: Paper-pencil simulations; multiple-choice questions; audio-visual simulations; live interviews

Group B: Multiple-choice questions; paper-pencil simulations; live interviews; audiovisual simulations

Group C: Audiovisual simulations; live interviews; multiple-choice questions; paper-pencil simulations

Group D: Live interviews, audiovisual simulations; paper-pencil simulations; multiple-choice questions

For the live interviews, subjects within each group interviewed the programmed patients in a randomly determined sequence.

The content for the live interviews was selected from five patient protocols. Non-professional actors were selected and trained to portray a specified role. The period of time required for orientation and training of each programmed patient equaled approximately four hours.

The studies were conducted at each site during a two-day period. Each subject experienced two evaluation methodologies the first day and the remaining two methodologies during the second day. Each live interview was audio recorded for subsequent analysis by members of the project committee. Subjects were told they were participating in an experimental study but were not informed that the purpose of the study was to assess interpersonal and communication skills.

Following the completion of the study, a different set of audio tapes was sent to each member of the project committee along with retraining materials, blank interaction observation forms, and instructions for analyzing the tapes. Members of the committee did not analyze audio recordings which were made at their own physician's assistant training program. A total of 295 audio recordings were analyzed. Since members of the project committee had participated in the two reliability studies, it was felt that this experience along with the retraining materials provided an adequate background for analysis of the audio recordings.

Findings

Table 6 displays the correlations among the five interpersonal skills variables as assessed by the live interviews. Efficiency does not correlate significantly with any of the other variables except Responsiveness to Patient Behaviors (Variable #5, .365). In fact, all interpersonal skills variables correlate more highly with Responsiveness to Patient Behaviors than with any other variable.

The explanation of the high correlations among Variables #2 through #5 lies in the distribution of behaviors across variables. For example, of the 8 behaviors included in Variable #2 (Non-Directiveness), 3 are also included in Variable #3 (Facilitation of Expression/Rapport-Support) and 4 are included in Variable #5 (Responsiveness to Patient Behaviors). One would hypothesize, then, that the correlation between Variables #2 and #5 would be greater than that between Variables #2 and #4. These hypothesized relationships are, in fact, observed in Table 6. The coefficient between Variables #2 and #5 equals .732, while the correlation between Variables #2 and #4 equals .459.

A similar pattern of relationships can be observed between Variables #3 (Facilitation of Expression/Rapport-Support and Variable #5 (Responsiveness). Of the 11 behaviors included in Variable #3, eight are also included in Variable #5. Thus, it is not surprising to find that the correlation between these two variables equals .905.

The distribution of interpersonal skills behaviors across the five scoring variables is displayed in Table 7. In interpreting the magnitude of relationships among the five interpersonal skills variables, it should be remembered that some are particularly inflated due to the overlap in content across variables. The correlations to which this caveat applies have been underlined in Table 6.

Table 6

Correlations Among Interpersonal Skills Variables

LIVE INTERVIEWS

	Efficiency 1	Non-Directiveness 2	Facilitation of Expression/Rapport-Support 3	Enhancing Patient Understanding 4	Responsiveness to Patient Behaviors 5
1	--				
2	.162	--			
3	.191	**.682	--		
4	.191	**.459	**.727	--	
5	**.356	**.732	**.905	**.778	--

**p < .01

df = 57

Table 7

Distribution of Interpersonal Skills
Behaviors Across Scoring Variables

		Variable				
		1	2	3	4	5
1.	Non-pertinent, open-ended questions	X				
2.	Pertinent, open-ended questions	X	X			
3.	Non-pertinent, closed-ended questions	X				
4.	Pertinent, closed-ended questions	X	X			
5.	Inappropriately leading question		X			
6.	Reflecting as a probe		X	X		X
7.	Repetitious questions	X				
8.	Continuous questions		X			
9.	Interrupts - abruptly shifts focus		X	X		X
10.	Encourages patient to talk		X	X		X
11.	Uses medical jargon			X	X	
12.	Follows-up cues	X				X
13.	Fails to follow-up cues	X				X
14.	Repeats question verbatim		X		X	X
15.	Clarifies question				X	
16.	Appropriately summarizes			X		
17.	Inappropriately summarizes			X		
18.	Expresses understanding			X		X
19.	Responds with inappropriate affect			X		X
20.	Reassures appropriately			X		X
21.	Reassures inappropriately			X		X
22.	Answers questions directly				X	X
23.	Answers questions evasively				X	X
24.	Gives appropriate advice/instructions				X	
25.	Gives inappropriate advice/instructions				X	
26.	Provides appropriate information				X	
27.	Provides inappropriate information				X	X
28.	Expresses empathy			X		X

It is interesting to note that Efficiency does not correlate significantly with Non-Directiveness, Facilitation of Expression/Rapport-Support, or Enhancing Patient Understanding. There is a modest correlation between Efficiency and Responsiveness to Patient Behaviors (.356, $p < .01$), but the correlation may be due solely to the fact that 2 of the 6 behaviors contained in the Efficiency variable (following and failing to followup patient cues) are also included in the Responsiveness variable. Thus, it appears that efficiency does not account for a significant proportion of the variance included in the other more affective-laden interpersonal skills variables.

Although the major thrust of this study was not the investigation of relationships among the 28 interpersonal skills behaviors, inter-correlations among the 28 variables as measured in live interviews, were calculated in order to determine the extent of their independence. Table 8 contains a summary of the best predictors for each of the 28 interpersonal skills behaviors as depicted from the live interviews. All coefficients are at or beyond the .05 level of confidence. As can be seen from this table, even when an effort is made to focus on best predictors, the relationships among these variables are quite modest. Only Variables #2, #9, #16 and #18 show correlation coefficients above .50. Thus, it seems reasonable to conclude that interpersonal skills behaviors are only modestly correlated among one another. There are, however, some interesting relationships among behaviors which, while not surprisingly large in terms of magnitude, are logically consistent. Some examples of such relationships are as follows:

a. The more interviewers express empathy to patients, the more they express understanding of what the patient has said and reassure the patient appropriately. Conversely, the more interviewers express empathy, the less they inappropriately interrupt patients

or abruptly shift focus and answer patient questions evasively.

b. There is a positive relationship between providing inappropriate information and answering patient questions evasively.

c. There is a positive relationship between providing appropriate information and the extent to which the interviewer reflects patient comments or questions as a probe and appropriately summarizes or synthesizes patient comments. There is a negative relationship between providing appropriate information and responding to patient comments with inappropriate affect.

d. There is a positive relationship between expressing understanding of what the patient has said and appropriately summarizing and synthesizing patient comments. There is a negative relationship, however, with expressing understanding and inappropriately interrupting patients or abruptly shifting focus.

e. The more interviewers use medical jargon, the less willing they are to clarify their own questions and comments to the patient.

f. The more interviewers encourage patients to continue talking, the more they answer questions directly and the less they respond to patients with inappropriate affect.

g. The more interviewers interrupt patients and abruptly shift focus, the less likely they are to appropriately summarize and synthesize patient comments and express understanding of what the patient has said.

h. The more repetitious questions the interviewer asks, the less likely he is to give appropriate advice and instructions.

i. The more interviewers use inappropriately leading questions, the more likely they are to inappropriately interrupt patients and express inappropriate affect in response to patient comments.

Table 8

Correlations Among 28 Interpersonal Skills Behaviors Depicted in Live Interviews ($p \leq .05$)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1																				-.36								
2														.27						-.33								
3																												
4		-.56												-.29											.35	.38		
5																	-.26		.41				-.34				-.30	
6				-.25																								
7																												
8																												
9					.40							-.33				-.61	-.26	-.54	.38		.34			-.43				-.38
10			.28																-.37			.39						
11												-.35			-.44						.33							
12																						-.34						
13				-.29																								
14											-.44																	
15									-.61									.54								.36		
16																												
17									-.54							.54						.39	-.35		-.40			.36
18										-.37																-.35		
19					.11																							.35
20	-.36		-.33																					-.36			
21																									-.36			
22										.39			-.34														.30	-.36
23						-.34															-.36							
24							-.33		-.43				-.33									-.36					-.31	
25																												
26																									-.31			
27						-.30														.35			-.36					
28																												

Since the major thrust of this project was to evaluate the effectiveness of several indirect testing techniques in predicting performance as assessed in a live interview setting, within and between-method correlation coefficients were calculated across the 5 interpersonal skills variables. These correlations are presented in Table 9. From the standpoint of within-method validity, the best predictor of each live interview variable should be the corresponding variable on the indirect technique. The indirect technique which produces the largest number of such corresponding correlations can be thought of as the most valid.

For the interactive audiovisual simulations, the live interview Efficiency score was best predicted by the corresponding audiovisual simulation Efficiency variable. For the live interview variable, Enhancing Patient Understanding, the corresponding audiovisual simulation variable also yielded the best prediction. The same was found to be true in relation to the variable, Responsiveness to Patient Behaviors, on the live interview and the corresponding audiovisual simulation variable. Thus, for the interactive audiovisual simulations, the appropriate pattern of validity coefficients was present in 3 of the 5 interpersonal skills variables.

For the paper-pencil simulations, there was a correspondence between live interview variables and the corresponding variables on paper-pencil simulations in only 1 out of 5 instances: Facilitation of Expression/Rapport-Support.

For the multiple-choice questions, the correspondence between scoring variables was observed in only 1 instance, and involved the Responsiveness to Patient Behaviors variable.

Thus, in relation to validity coefficients, interactive audiovisual simulations yielded the largest number (3) as compared with paper-pencil

Table 9

Simple Correlations Among IPS Variables Within and Between Measurement Methods*

	VARIABLE	METHOD: MCQ 1	MCQ 2	MCQ 3	MCQ 4	MCQ 5	PMP 1	PMP 2	PMP 3	PMP 4	PMP 5	AVSIM 1	AVSIM 2	AVSIM 3	AVSIM 4	AVSIM 5	LIVE INTERVIEW 1	LIVE INTERVIEW 2	LIVE INTERVIEW 3	LIVE INTERVIEW 4	LIVE INTERVIEW 5
MCQ	1. Efficiency		.066	.209	.111	.309	.220	.428	.232	-.179	.303	.158	.230	.116	-.205	.055	-.013	.216	.183	.148	.214
	2. Non-Directiveness			.475	.329	.839	.047	.370	.278	.108	.270	.289	.077	.254	.037	.314	-.114	.325	.252	.141	.181
	3. Facilitation of Expression/ Rapport-Support				.284	.713	.243	.432	.362	.060	.271	.323	.044	.183	.088	.221	-.058	.287	.224	.078	.230
	4. Enhancing Patient Understanding					.558	.151	.333	.373	.198	.393	.283	.214	.457	.124	.810	-.048	.288	.388	.145	.300
	5. Responsiveness to Patient Behaviors						.129	.406	.308	.094	.370	.282	.082	.248	.178	.348	-.088	.400	.331	.181	.317
PMP	1. Efficiency							.337	.402	-.318	.583	.301	.270	.261	.015	.282	-.049	.302	.222	.205	.230
	2. Non-Directiveness								.854	-.048	.707	.380	.537	.473	-.205	.358	-.160	.253	.228	.108	.234
	3. Facilitation of Expression/ Rapport-Support									-.222	.878	.288	.458	.574	-.118	.407	-.253	.284	.288	.228	.278
	4. Enhancing Patient Understanding										-.094	.090	-.098	-.078	.291	-.017	.121	.028	.057	.018	.048
	5. Responsiveness to Patient Behaviors											.298	.428	.478	.007	.391	-.244	.345	.274	.183	.281
AVSIM	1. Efficiency												.223	.335	-.188	.532	.210	.383	.360	.067	.308
	2. Non-Directiveness													.858	-.133	.387	-.008	.088	.104	.138	.108
	3. Facilitation of Expression/ Rapport-Support														.088	.814	-.028	.300	.332	.301	.330
	4. Enhancing Patient Understanding															.282	.182	.158	.251	.372	.281
	5. Responsiveness to Patient Behaviors																.019	.398	.450	.335	.426
LIVE INTERVIEW	1. Efficiency																	.182	.191	.181	.358
	2. Non-Directiveness																		.682	.458	.734
	3. Facilitation of Expression/ Rapport-Support																			.727	.805
	4. Enhancing Patient Understanding																				.778
	5. Responsiveness to Patient Behaviors																				

*The units of analysis were compilations of ranks, transformed to standard scores, of subsets of the 28 behavioral categories. (See Walker & Lev, Statistical Inference. New York: Holt, Rinehart & Winston, 1953, p. 424).

simulations and multiple-choice questions which yielded 1 each. The coefficients which provide between-method indications of validity have been circled on Table 9.

Table 10 shows the simple correlations between the 5 live interview variables and the corresponding variables for each indirect testing technique. As can be seen from this table, none of the Efficiency scores on the indirect techniques correlate significantly with the Efficiency score in live interview settings. The correlation coefficient of .21 between Efficiency in the live interview setting and Efficiency as measured by the interactive audiovisual simulations comes closest to reaching statistical significance at the .05 level of confidence (r of .25). In relation to Non-Directiveness, the corresponding variable as measured by multiple-choice questions yielded the largest correlation coefficient (.325). This coefficient is significant beyond the .01 level of confidence. In terms of Facilitation of Expression/Rapport-Support, Enhancing Patient Understanding, and Responsiveness to Patient Behaviors, the corresponding variables as assessed by interactive audiovisual simulations yielded the largest correlation coefficients. These coefficients equaled .332, .372, and .426 respectively and are all significant beyond the .01 level of confidence.

In summary, Table 9 indicates that 4 of the 5 interpersonal skills variables can be best predicted by performance as assessed by interactive audiovisual simulations. In one instance, (Non-Directiveness) this prediction was best accomplished through the use of multiple-choice questions. It should be again pointed out that the prediction of each interpersonal skills variable was limited to the corresponding variable as measured by each of the indirect testing techniques.

Table 10

Simple Correlations Between
Live Interview Variables
and Testing Techniques

	Efficiency	Non-Directiveness	Facilitation/Rapport-Support	Enhancing Understanding	Responsiveness
	1	2	3	4	5
Multiple-Choice Questions	-.013	**.325	.224	.145	*.317
Paper-pencil Simulations	-.049	*.253	*.289	.018	*.261
Audiovisual Simulations	.210	.086	**.332	**.372	**.426

*p = .05

**p < .01

df = 57

In order to estimate which variables across testing techniques would be the best predictors of each live interview variable, a step-wise multiple regression analysis was performed. The results are presented in Table 11. There was no scoring variable or indirect testing technique which would predict Efficiency at the .05 level of confidence. For this reason, it was not possible to obtain a regression solution for this variable. The other results of the regression analysis indicate that Non-Directiveness can be best predicted by combining the Responsiveness scores as measured by multiple-choice questions and interactive audiovisual simulations. Under these circumstances, the multiple correlation coefficient equals .485. Similarly, it was found that the best predictor of Facilitation of Expression/Rapport-Support is also the Responsiveness score as assessed by interactive audiovisual simulations, resulting in a correlation coefficient which equals .450. The best predictor of Enhancing Patient Understanding is the corresponding variable as assessed by interactive audiovisual simulations plus Facilitation of Expression/Rapport-Support also as measured by interactive audiovisual simulations. When these two predictors are combined, they result in a multiple correlation coefficient equal to .462. Finally, Responsiveness to Patient Behaviors is best predicted by the corresponding variable as assessed by interactive audiovisual simulations. This correlation coefficient equals .425.

Perhaps the most striking aspect of the multiple regression analysis is the fact that of the 6 predictors identified, 5 are obtained from assessment by the Interactive Audiovisual Simulation System. Moreover, of the 5 scoring variables into which behaviors were classified, Responsiveness to Patient Behaviors appears to be the most potent predictor of other variables as well. An explanation for this phenomenon can be found in the

Table 11

Best Predictors of Variables in Live Interview

Variable in Live Interview**	Predictors	Multiple Correlation*	Simple Correlation*
2. Non-Directiveness	MCQ Responsiveness	.400	.400
	AVS Responsiveness	.485	.398
3. Facilitation of Expression/Rapport - Support	AVS Responsiveness	.450	.450
4. Enhancing Patient Understanding	AVS Enhancing Understanding	.371	.371
	AVS Facilitation/Rapport	.462	.332
5. Responsiveness to Patient Behaviors	AVS Responsiveness	.425	.425

*all coefficients are singificant beyond .01 level of confidence
**F-level was not sufficient for the Efficiency variable to permit a solution

fact that the Responsiveness variable contains 13 of the 28 categories of interpersonal skills behaviors.

The fact that Facilitation of Expression/Rapport-Support is best predicted by Responsiveness to Patient Behaviors (measured by interactive audiovisual simulations), may be explained by the large amount of overlap in content between these two variables and the possibility that there is significant error variance in the behaviors which are included in the Facilitation variable but not in the Responsiveness variable.

A similar phenomenon may also be present with respect to the Non-Directiveness variable in that the non-overlapping behaviors may have little variance or considerable error variance associated with them. If such were the case, Responsiveness to Patient Behaviors could be expected to be the best predictor of Non-Directiveness.

Conclusions

The data presented above suggests that:

a. There is greater evidence for the convergent and divergent validity of interactive audiovisual simulations than for multiple-choice questions or paper-pencil simulations.

b. In predicting interpersonal skills behaviors as measured in a live interview setting, the corresponding variables are best predicted in 4 out of 5 instances by interactive audiovisual simulations when the basis for prediction is a simple correlation coefficient. In predicting the remaining variable (Non-Directiveness) with a corresponding variable from an indirect testing technique, the use of multiple-choice questions yields the largest correlation coefficient.

c. In predicting live interview variables across testing techniques by means of multiple regression, interactive audiovisual simulation is a significant predictor of all variables which were included in the analysis. In addition to interactive audiovisual simulation, Responsiveness as measured by multiple-choice questions emerged as the only significant predictor variable from the other two indirect testing techniques investigated in this study.

It seems reasonable to conclude that of the three indirect testing techniques investigated in this study, interactive audiovisual simulations yield the most potent predictions of performance in a live interactive setting. The magnitude of these correlations (.40 to .45) suggests that additional work should be undertaken in order to further refine this innovative technique and to determine the maximum level of prediction possible. In view of the fact that the interactive audiovisual simulations and the corresponding Interactive Audiovisual Simulation System are in prototype stages of conceptualization and development, the magnitude of the correlation coefficients is particularly encouraging. It suggests that further refinement of this evaluation methodology could significantly enhance its contribution as an approach to the assessment of the interpersonal skills of health professionals.

AUDIT VALIDATION STUDY

Introduction

National examinations for licensure and certification in the health professions have traditionally been used as one of the components in evaluating qualifications for practice. The major purpose of these examinations is to identify candidates who are not yet ready to assume responsibility for practice in a given health profession. In essence, such examinations serve as a fail-safe mechanism for protecting public safety and the integrity of the profession itself. Since it is generally assumed that the quality of a profession is only as good as the skills of the individuals who practice it, the maintenance of high standards of practice is essential.

Because of the role that licensing and certifying examinations play in this process, their structure and content need to be continually evaluated in order to ensure that they are measuring knowledge and skills relevant to practice in the profession. It is assumed that if a certifying examination does, indeed, assess relevant aspects of competence, then performance on the examination should have a positive relationship to subsequent performance in practice.

It is doubtful that anyone would argue the merits of the above assumption. What does become a point of some debate is the nature of the specific relationships that should be expected, and the magnitude of these relationships. If one returns, for a moment, to the role of certifying examinations, one finds that they are used only as one component in an overall evaluation process. The reason that such examinations are not used as the sole component is twofold. First, a single evaluation program cannot be expected to measure all of the areas of knowledge and skill required for competent practice. Secondly, no measure of

competence, be it grades, formal examinations, faculty ratings of performance or actual observations of performance yields perfect predictions of subsequent behavior in a practice setting. Thus, using several sources of data in evaluating qualifications for practice serves to enhance the accuracy of the decision to grant or withhold certification. By ensuring that there is a positive relationship between performance on these input measures and subsequent performance in practice, we help ensure the validity of our evaluation procedures. Certifying examinations will not be able to assess all relevant components of competence, and for those components they do assess, they will not do so perfectly.

These observations are in no way meant to serve as an apologia for certifying examinations. Rather they are meant to place these examinations into perspective so that reasonable expectations can be specified and tested.

The Physician's Assistant Certifying Examination has undergone a number of validation studies during its first three years of implementation. These studies have shown significant positive relationships between scores on various components of the Certifying Examination and faculty ratings of these components of competence. While the magnitude of these relationships has not exceeded .40, the patterns of significant correlation coefficients indicate stronger relationships where one would expect to find them and only chance level relationships where logically one would not expect to find a correlation between examination scores and ratings of clinical competence. In addition, physician's assistants with up to two years of clinical experience following completion of an educational program achieve significantly higher examination scores than candidates who sit for the examination before completing their training program. This finding suggests that the examination is measuring knowledge and skills that are relevant to practice since candidates with practice experience achieve significantly higher scores on the examination.

An important relationship that has not yet been explored, however, is the extent to which performance on the Certifying Examination is predictive of certain aspects of day-to-day performance in an actual clinical practice. In order to investigate this relationship, the National Board of Medical Examiners, with the endorsement of the American Academy of Physicians' Assistants, embarked upon a study to obtain measures of performance in an actual practice setting so that these could be correlated with scores on the Certifying Examination.

Obviously, the assessment of performance in a practice setting is a very complex endeavor. The assessment procedure employed must not only yield valid and reliable observations of performance, but it must focus on relevant aspects of practice behavior in a manner that is unobtrusive and cost effective.

How does one determine whether an adequate history and physical examination were performed and the appropriate laboratory and diagnostic studies ordered? How does one evaluate the physician assistant's skill in interpreting these clinical findings? How does one determine whether the P.A. uses appropriate judgment in seeking consultations and referral? How can one assess whether appropriate management and treatment was instituted and adequate instructions provided to patients?

If these are the major components of competence we wish to use as criteria for evaluating practice performance, we need to consider the evaluation methodologies that are available for the collection of these criterion data. Three methodologies exist: a) direct observation of P.A. performance by a trained observer; b) employer ratings of P.A. performance; and c) reviews of office medical records.

Each of these methodologies poses certain limitations and problems for the evaluation of practice performance. For example, direct observation of performance involves significant amounts of time and expense. Moreover, the

presence of an observer requires the patient's consent and can bias the P.A.'s behavior. Employer ratings of performance can be more time and cost effective than the previous methodology cited, but are subject to the same problems generally associated with the use of rating scales -- such as, a tendency to be very lenient in the ratings of performance, and to rate an individual in the same way across the various competency components under consideration. Because of the subjectivity of these ratings, they may not provide accurate evaluations of clinical performance. Finally, the review of office medical records provides a reasonably objective basis for documenting behavior, although it does not provide a mechanism for evaluating the physician assistant's skill in interpreting clinical data or judgment in seeking advice and consultation. Moreover, the medical record may not fully document all pertinent aspects of the evaluation and management of a given patient. Thus, in order to use the medical record as a basis for documenting performance, one must assume that the record provides a reasonably accurate documentation of clinical behavior.

When cost effectiveness and objectivity of the evaluation process were considered for each of the methodologies noted above, it appeared that the review of office medical records could provide the best source of data regarding practice performance.

The review of office medical records is based upon the identification of specific medical conditions which serve as the guide for selecting appropriate charts for review. The conditions selected should include those which are seen in the practice setting with sufficient frequency to provide an adequate sample of medical records. Since the Certifying Examination focuses on primary care, it was felt that the evaluation of practice performance should be based upon physician's assistants who are practicing in primary care settings.

Procedures

Having selected medical record audits as the methodology to be used in gathering data regarding practice performance, several objectives needed to be accomplished. The medical conditions and associated audit criteria had to be identified; these criteria needed to be pilot-tested in order to ensure their appropriateness and clarity; individuals who would be responsible for auditing the office medical records had to be identified and trained; and procedures needed to be established to identify appropriate medical records within each practice and to abstract the data contained in each.

In order to maximize participation in this study, an endorsement was obtained from the American Academy of Physicians' Assistants. In addition, the Executive Committee of the Academy was asked to nominate physicians and physician's assistants who could serve on the Advisory Committee. The role of the Advisory Committee was to identify the specific medical conditions to be used in this study and to specify the audit criteria for each.

A ten-member Advisory Committee was appointed consisting of practicing physician's assistants, primary care physicians who employed physician's assistants, and physicians involved in the training of physician's assistants. At its first meeting, the Committee was asked to identify a number of medical conditions which they felt were representative of those seen in primary care practice settings. The Committee identified the following twelve conditions:

a. angina - initial and follow-up evaluations
b. asthma - initial and follow-up evaluations
c. birth control pills - initial and follow-up evaluations
d. diabetes mellitus - initial and follow-up evaluations
e. headache
f. health maintenance
g. hypertension - initial and follow-up evaluations

h. acute purulent otitis media

i. tonsillopharyngitis

j. urinary tract infection

k. vaginitis

l. venereal disease

Each member of the Advisory Committee was assigned the responsibility of developing audit criteria for several of the medical conditions so that these criteria could be reviewed and refined by the entire Committee at subsequent meetings. Committee members were advised that the audit criteria should consist of procedures felt to be essential to the diagnostic work-up and management of patients with the specific medical condition. Essential procedures were felt to be those which were necessary to establish the correct diagnosis and to resolve the patient's medical problem.

For each medical condition, historical, physical examination, laboratory data, and management plans that should be found in the medical records of patients being seen for evaluation were specified. (A sample of the audit criteria specified for two of the medical conditions is provided in Appendix D.) In addition to the criteria themselves, examples of acceptable and unacceptable chart entries were also specified so that record abstractors could determine whether a specific chart entry satisfied a given criterion. Both the audit criteria and the examples of chart entries were specified on the basis of consensus reached by the Advisory Committee after careful discussion and pilot use of the criteria with a sample of medical records. (The pilot use of the criteria was carried out by the project staff, using anonymous medical records that had been provided by individual members of the Advisory Committee.)

An abstracting form was prepared for each set of audit criteria. The form was designed so that record abstractors could record whether or not a given medical record satisifed each criterion specified for a given medical condition. (Several sample abstracting forms and instructions for abstractors are provided in Appendix E.) For each criterion the abstractor was required to make one of four decisions: "Yes," if the chart entry satisfied the criterion; "No," if the chart entry did not satisfy the criterion; "UC," if the abstractor was uncertain as to whether or not the chart entry met the criterion; "NA," if the criterion was not applicable for the specific patient whose medical record was being audited. (Such instances were often related to age and sex variables, or the presence or absence of specific diagnosis findings.)

In order to obtain an adequate sample of medical records for each participant in the study as well as to facilitate the retrieval of these records, a practice log form was developed and sent to each physician's assistant. Each participant was instructed to use the log for two months during which time he/she would record the names or identification numbers of patients who were seen for one or more of the medical conditions included in this study. When the spaces in the log form were filled, the physician's assistant was asked to tally the number of patients who had been seen for each medical condition and forward this summary to the project staff. The log form itself was retained in the practice for use as a guide in subsequent record retrieval.

The privacy of patients whose medical records would be audited was protected by withholding from the project staff any information concerning patient identity. In order to ensure that this procedure was followed, each participating physician's assistant was asked to review the abstracting

forms before they were forwarded to the National Board to determine that no patient identifications had been recorded. In addition, all abstractors were cautioned about the need to maintain strict anonymity of all patient data.

Abstractor Training and Reliability Study

Seven record abstractors were recruited from around the country and brought to Philadelphia for a one-week training program. All abstractors were registered record room technicians or administrators who had previous experience in auditing medical records.

During the one-week training program, the audit criteria and examples of chart entries for each medical condition were presented and reviewed in detail by a physician (and in one instance, a physician's assistant). Abstractors were encouraged to ask questions so that ambiguities could be clarified and an optimum level of understanding achieved. Following the discussion of each set of audit criteria, the abstractors were given several medical records and asked to audit each using the appropriate abstracting form. This provided abstractors with the opportunity to practice with each set of criteria and to identify areas requiring further clarification. This process was repeated for each medical condition.

At the conclusion of the training program, a reliability study was conducted in order to determine whether the abstractors were functioning with sufficient accuracy. Since time would not permit the abstractors to complete a set of medical records for all twelve medical conditions included in the study, eight medical conditions were selected. During the reliability study, the abstractors were instructed not to speak with one another or to discuss either the medical records or audit criteria. A member of the project staff

remained in the room with the abstractors during the entire study in order to ensure that the rules of conduct were followed.

In order to obtain an estimate of reliability, the coefficient of inter-abstractor agreement proposed by Maxwell and Pilliner was used. This coefficient is the equivalent of an intra-class correlation, but has additional advantages in that it is based on individual item responses rather than on an overall score. Ideally, a coefficient of abstractor agreement should be greater than zero as the amount of agreement departs from chance level, and should fall below zero if one abstractor records his judgment regarding a criterion in the opposite direction from another abstractor using the same criterion. The Maxwell and Pilliner correlation coefficient equals 1.00 only if there is complete agreement between two abstractors, and equals 0 if the amount of agreement between abstractors could be expected by chance alone. The coefficient equals -1.00 if there is complete disagreement between two abstractors using the same criteria in relation to the same medical records.

Table 12 presents the correlation coefficients between indicated pairs of abstractors for eight medical conditions. The overall correlation coefficients range from .63 to .79 with a mean of .72. These findings are consistent with levels of inter-abstractor agreement observed in similar audit projects at the National Board.

Following completion of the training program, each abstractor was assigned a group of practices in a specified region of the country. All participants had been advised that a project abstractor would be visiting their practice within the next several months. Each participant was then advised of the name of the abstractor who would visit his/her practice, and each abstractor was given a letter of introduction to serve as a means of appropriate identification.

Table 12

ABSTRACTOR RELIABILITY ESTIMATES

Audit Validation Study

Abstractor Pairs	Data Base							
	Asthma (02)	BCP (03)	Hlth. Maint. (06)	OM (08)	Tonsil-lophar. (09)	Vagin. (11)	V.D. (12)	Asth F/ (15
1 vs 2	.55	.73	.90	.81	.86	.94	.74	.8
1 vs 3	.70	.69	.73	.90	.67	1.00	.67	.6
1 vs 4	.47	.78	.93	.75	.86	.81	1.00	.6
1 vs 5	.70	.69	.92	.61	.82	.67	.76	.8
1 vs 6	.41	.62	.71	.81	.72	.87	.64	.6
1 vs 7	.92	.75	.80	.75	.89	.81	.86	.6
2 vs 3	.85	.61	.68	.70	.72	.94	.87	.5
2 vs 4	.60	.58	.82	.56	.81	.74	.78	.5
2 vs 5	.85	.82	.87	.61	.87	.73	.74	.6
2 vs 6	.42	.59	.70	.61	.77	.94	.57	.5
2 vs 7	.60	.70	.84	.65	.84	.88	.65	.5
3 vs 4	.72	.74	.72	.85	.71	.81	.69	.4
3 vs 5	1.00	.79	.78	.70	.75	.67	.68	.8.
3 vs 6	.55	.41	.73	.90	.87	.87	.69	.6
3 vs 7	.76	.73	.62	.85	.71	.81	.69	.3
4 vs 5	.72	.62	.85	.65	.75	.61	.77	.77
4 vs 6	.56	.60	.69	.85	.68	.81	.69	1.00
4 vs 7	.53	.66	.72	.80	.94	.63	.90	.58
5 vs 6	.55	.49	.73	.71	.83	.67	.46	.77
5 vs 7	.76	.66	.77	.85	.88	.62	.82	.57
6 vs 7	.33	.45	.64	.85	.69	.81	.66	.57
Overall	.63	.64	.77	.75	.79	.79	.72	.65
No. of records abstracted	2	3	3	4	4	3	4	2
No. of items/record	25	22	39	12	13	18	32	11
Total No. of items	50	66	117	48	52	54	128	22

Sample Selection

A stratified random sample of 800 physician's assistants who took the 1974 Certifying Examination was selected for this study. The variables used to stratify the sample included:

a. type of training - whether formally or informally trained;

b. type of formal training - that is, physician's assistant or Medex program; and

c. composite examination score

Accordingly, 800 letters of invitation were sent including a description of the purpose of the study and the procedures which would be used to conduct it. In all instances, the American Academy of Physicians' Assistants was asked to supply current addresses for those individuals who were members of the Academy. Of the 800 physician's assistants invited to participate in the study, 222 indicated that they did not wish to do so. Sixty-one physician's assistants could not be located by mail or had moved without leaving a forwarding address. In these instances, letters of invitation or follow-up letters were returned to the National Board and further attempts to obtain a current mailing address were unsuccessful. Two hundred eighty-seven physician's assistants did not respond to either the initial letter of invitation or the follow-up letter. Since these letters were not returned by the post office, it was assumed that they had been received. Finally, 230 physician's assistants indicated their willingness to participate in the study. Of that number, 96 dropped from the study for a variety of reasons, leaving a final sample of 134 physician's assistants whose records were audited. These aspects of the sample size are summarized in Table 13.

Table 13

Summary of Sample Size
Audit Validation Study

Category	Number
Letter of Invitation Sent	800
Could not be Located	61
Did not Reply	287
Unwilling to Participate	222
Willing to participate	230
Dropped during Study	96
FINAL SAMPLE SIZE	134

The ninety-six participants who dropped out of the study after indicating their willingness to participate did so for a variety of reasons. After being advised of the medical conditions included in this study, twenty-one physician's assistants indicated that they were not in primary care practice. In eighteen instances, physician's assistants could not participate further because the necessary medical records were not available. Eleven physician's assistants reported that at present they were not in active practice, and ten indicated that they were in the process of changing practices. Nine physician's assistants were located in remote sites such as Alaska, Hawaii, and Germany, and it was determined that it would not be cost effective for these individuals to participate in the study. Five physician's assistants subsequently reported that they were unemployed and 18 subsequently declined to participate further in the study. These reasons for dropping out of the study are summarized in Table 14.

Table 14

Reasons for Dropping from Study

Reason	Number
In specialty practice	21
Records not available	18
Declined to participate further	17
Not in active practice	11
In process of relocating	10
In a remote site (Alaska, Hawaii, etc.)	9
Unemployed	5
Miscellaneous	5
TOTAL	96

Of the 134 physician's assistants who participated in all aspects of the audit validation study, 87 per cent were graduates of formal educational programs and 13 per cent had qualified for the Certifying Examination on the basis of work experience. Of those trained in formal educational programs, 71 per cent were trained in physician's assistant programs and 18 per cent in Medex programs. (Data were not available for 11 per cent of the participants.) In terms of the examination performance of this group of participants, their mean written examination score equaled 508 with a standard deviation of 76. The 1974 examination statistics for the total examinee population were as follows: the mean written score equaled 492, while the standard deviation equaled 83. The characteristics of the study sample in relation to the total examinee population are summarized in Table 15.

Table 15

Representativeness of Study Sample
to Total Examinee Population

Factor	Examinee Population		Study Sample	
Demographic	N	%	N	%
Formally Trained	1118	91	110	86
Physician's Assistants	922	75	95	71
Medex	196	16	21	16
Informally Trained	115	09	18	13
Exam Statistics	Written	PAS*	Written	PAS*
Composite Mean	492	485	508	485
Standard Deviation	83	112	76	110

*Physical Examination skill assessment

It appears from the data concerning demographic characteristics and examination statistics of the study sample that it is reasonably representative of the total examinee population for the 1974 Physician's Assistant Certifying Examination. However, it does represent only 10% of the number of individuals who sat as candidates for that examination, and represents 17% of the initial stratified random sample of 800 physician's assistants who were invited to participate in the study. The fact that 509 individuals either did not reply to our initial and follow-up letters or indicated that they were unwilling to participate is likely due in large measure to the sensitive nature of medical record audits. Although all participants were assured of the confidentiality of the data and the fact that patient identities would remain anonymous, there may well have been concern about allowing sensitive data such as that contained in office medical records to be reviewed by individuals who were not health care providers in the practice. It is also likely that issues related to

malpractice suits, third party reimbursement, and physician relicensure and recertification were contributing factors in decisions not to participate in this study.

Data Collection

As was mentioned in an earlier section, each abstractor was assigned a group of practices to visit during the three-month period available for this activity. Abstractors were instructed to contact participants by telephone and arrange an appointment to visit the practice. The physician's assistant was informed of the number of medical records to be abstracted for each medical condition and was asked to have these records assembled for the abstractor by the time of her appointment.

Since each physician's assistant had retained his/her own practice log form, it was felt that this form could serve as the basis for retrieving appropriate charts. However, it was more often the case that when the abstractor arrived at the practice, she found that the indicated records had not been pulled. Thus, a half day or more was spent attempting to retrieve appropriate charts utilizing record systems that were often not well organized. Because of these circumstances, abstractors were often not able to review as many medical records as would have otherwise been possible. Moreover, in some instances the format of the records themselves resulted in longer periods of time being required for the audit. As soon as the data collection at each practice had been completed, the abstracting forms were forwarded to the National Board for processing.

A frequency distribution of the number of charts per medical condition for each study participant is provided in Table 16. Audit data from a total of 2954 medical records were available for analysis.[1] It is interesting to

[1]This total represents the records of 116 participants. The records of 18 physician's assistants had to be dropped from the analysis because of questions concerning the reliability of the data.

note that there is considerable variation in the number of medical records audited across the various medical conditions. If one can assume that the number of medical records audited for each medical condition is a reasonable approximation of the relative frequency with which P.A.'s see patients with these medical conditions, the frequency distribution suggests that physician's assistants spend more time managing certain medical conditions and very little time managing others. For example, the number of records of patients with tonsillopharyngitis, otitis media, and hypertension as well as patients being seen for health maintenance is far greater than the number of records of patients with asthma, angina, or those seeking birth control devices. Although there is a fair degree of overlap among participants in terms of the types of medical records available for audit, there are a number of medical conditions for which there are little available audit data. Moreover, in a number of instances there were audit data from only one medical record in relation to a medical condition for a given participant. Because it has been our experience with other medical record projects that the performance of a given health professional varies dramatically across patients being seen for the same medical condition, it was felt that the availability of only one medical record for a given medical condition would not constitute a sufficiently representative sample of the behavior of a physician's assistant. For this reason, when audit data from only one medical record were available for a given condition, they were not included in the statistical analyses. (These cases have also been dropped from the frequency distribution).

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
101				2		4							3	10				19
103						8			3				2	3				16
105					3	4			11									18
106				2			8						6	5			8	27
109								5	5					3				13
111	3	3			2	2		2	2	2	2	2		2	2			24
114									12	8			4	15				39
116					6	7	7	5	9	2		5						41
117		6																6
118				2			2	5	5	2				5				21
119					2	6		2	5	5	5			5				30
120	4					2		2	8					2				18

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
121					2	2		2	3	3			3	4	2		3	24
122					3	4	3		8		2	3		2				25
123						3		2	7				2	3				17
125						4		5	3		2			2	2			18
126														2				2
127								7	7	6		5						25
129					2			2	10	3				2				19
130			2		3	3		3	4		2		3	3		2	2	27
131		2						5	26	7	9	2						51
132					2				6			4				2		14
133		3			2				21			3						29
134						3	2	3	5	2		2		3				20
135					2	3		8	12	2	2		11	10				50

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
136								5	12				2	11				30
137								3	8	2				2				15
142					3			5	10			4		3	2			27
144						2		14	12	4		2						34
146					2	3		2	2	3			2	2	2		2	20
147	2				2	2		7	8				3	5				29
148						3		5	17	2				3				30
150					2	2	2	2	8				3	7			3	29
153			2		2	2	2	3	4		2	2	4	4		2		29
154								3	9			2						14
155													5	16			11	32
157					3			2	3	3	3	2		2	3			21

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
158		2				3		5	10			2		5				27
159		4						5	8									17
161					3	3	3	3	3	3		3	3	3				27
164	4				2		5			3								14
166	2			2		2		2	3	3		4	3	3	2		2	28
168					2	10			11	8	3							34
170					2	2	3	3	5	3	3		3	3		4	2	33
173					5			2	4	3	2		2					18
175						2		10	14	3	3					2		34

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
176					4	4		5	10		3							26
177	2													2			2	6
178							3	3	9	5	2			2				24
179					3	2		9	10	2			2	2		2		32
180					3		2	4	10	4	2			2				27
182						14		6	6					2				28
187							2	5	12	4				2				25
190						4		2	8					7			2	23
192									5					2				7
194			3		2	3	3	4		3				2				20
197	4			4		8	5		2					3			2	28
199					3		2	2	3	2	2		2	3	2			21

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
200							2							2				4
201					4				11	4		4		3				26
202						6	3						5	5				19
203					4	6		8	11				5	7			6	47
204					3			3	5	4	2							17
209		4			2		3		4	4	3	2						22
210			2		2	2		2	2		2		2		2		2	18
211						9		5	5									19
212						2			4	2			5	10	3	2	2	30
213		2						4	3	3	3		7	7	4		3	36
214		2			3			2	14	5	4							30
216	2					4		3	4	4	2							19
218								4	7		3					2		16
219						10		7	16					7				40

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
222								4	24									28
223					3			4	18	3								28
224						4		5	6				3	5				23
226					2	4			2	2								10
227						2	2							28				32
228								5	16				5	2		5		33
229						6		3	8				4	4	2		4	31
230						4		2	9				2	10				27
231				2		3	3		8	2			2	5				25
232					2	3	2	4	4	2	2	3		4				26
233									4	4			3	11	4			26
234						2	2	3	3	2	3				3	5		23
236										3	30	6						39

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
237					3				8	2		2						15
238					4	3	2	2	4		3	2		5		2		27
239		2			3	3		6	10	3	10							37
241		3			2		2	4	4	3	2							20
245						4		5	9			3			5			26
248						3		5	10	2				2	2			24
250						5		3	7	3	3	4						25
251					4		2	14	17	3				4				44
252		3			6			3	10			2						24
255								17	5						3			25
256	5			7	6	3	12	7	15	8	8		4	7				82
257	2		2		2	2	2	2	3	3	2		3	3	2	2	2	32

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
259	2				2	10			5									19
260						12			2	2			2	4			3	25
261				5			3						6	7			3	24
262								6	16	3	2							27
264					4	2		3	6	2		3						20
265					2		2	3	3	3				8				21
266		2			2	2	2	2	2	2	2		2	2		2	3	25
267					3	2	2	3	5					4	2		3	24
272			4		4			3	3	2	3	4		2		5		30
273						7			8	2								17
274						6			8		4			2				20

Table 16

Number of Abstracts Per PA Per Data Base

PA#	Angina	Asthma	Birth Control Pills	Diabetes	Headache	Health Maintenance for Adults	Diastolic Hypertension	Otitis Media	Tonsillopharyngitis	Urinary Tract Infection	Vaginitis	Venereal Disease	Diabetes Follow-up	Diastolic Hypertension Follow-up	Asthma Follow-up	Birth Control Pills Follow-up	Angina Follow-up	TOTAL
275					2			4	6	2								14
276								5	6	2		2						15
277		2			3			3	9	3		3		11				34
278						7		9	20	3				19				58
279				3	2						2	2	2	4	2	2	2	21
280			2		3	5		2	2	2	5			5				26
281						4		2	4	4				4				18
282	4	2			3	4	5	3	7	4		3		2		2		39
116	36	42	17	29	157	277	105	363	800	209	149	92	130	384	51	43	70	2954

It will also be noted from Table 16 that in instances where more than one medical record is available, the actual number of records for a given participant may vary from two to as many as thirty for a given medical condition. Because of the analytical problems posed by this variation in the number of records available for each participant, all scores were standardized prior to the statistical analyses as described in subsequent sections. However, even given the standardization process, one needs to be aware of the fact that the number of subjects included in the various statistical analyses performed on these data is not uniform.

Scoring and Measurement Issues

The audit data for all medical records were treated statistically by calculating a compliance score for each using the following formula:

$$\frac{\text{Sum of "Yes" entries}}{\text{Sum of "Yes" + Sum of "No" entries}} \times 100$$

where,

"Yes" indicates the audit criterion was met
"No" indicates the criterion was not met

Because of the variability in the number of records per medical condition across participants, all compliance scores were standardized prior to further analysis. One of the most fundamental measurement issues related to the analysis of the data from this study involves the nature of the predictive instrument (i.e., the Certifying Examination) and the criterion measure (i.e., audit data). While the data on the predictive instrument was the same for all subjects in this study, the data on the criterion measure is probably not identical for any two subjects. In addition, it should be remembered that the Certifying Examination is designed to assess aspects of knowledge and skills which cannot be assessed directly by an audit of medical records.

In order to provide a sharper focus for the analysis of data between the predictive instrument and the criterion measure, an attempt was made to classify the data from both sources in relation to what appeared to be logical components for analysis. In relation to the Certifying Examination, the test items were classified into the following components of competence:

a. data gathering on patient management problems (PMP's)

b. management-treatment on patient management problems (PMP's)

c. interpretation of clinical findings on multiple-choice questions (MCQ's)

d. management-treatment on multiple-choice questions (MCQ's)

Within each of the components of competence listed above, a broad sample of medical content and medical conditions is included. In fact, it should be pointed out that the sampling of medical conditions included on the Certifying Examination is far broader than that which was attempted for the medical record audit. To the extent that we know from previous statistical analyses that performance on examination materials varies as a function of the medical content being sampled, it should be recognized that the medical content of the predictive and criterion measures could not be controlled in this study. Therefore the magnitude of relationships observed between performance on the Certifying Examination and the medical record audit needs to be viewed in this context.

In relation to the medical record audit, it was felt that the data were related to two components of competence:

a. skill in diagnostic data gathering; and

b. skill in patient management (including patient counseling and instruction).

However, in reviewing the number of audit criteria related to management for each condition, it was determined that the number of items was too small

to constitute the meaningful score. Thus, two audit scores were possible for each medical condition:

a. data gathering; and

b. composite (consisting of data gathering and management criteria).

These two scores were used to describe performance in relation to most of the medical conditions used in the record audit. In some instances, the number of data gathering criteria was too small to be treated as a separate scoring component. When this occurred, only a composite audit score was reported for the medical condition.

In addition to developing scores for each medical condition, clustered scores were developed for some medical conditions which were grouped together. Three groups of medical conditions were developed:

Cluster A: Initial evaluations for headache and hypertension

Cluster B: Initial evaluations for tonsillopharyngitis, otitis media and health maintenance

Cluster C: Initial evaluations for tonsillopharyngitis, otitis media, urinary tract infection and vaginitis

The identification of additional clusters was not possible because of the relatively infrequent availability of audit data for the remaining medical conditions.

Findings

In order to determine whether there was a correlation between compliance with audit criteria and performance on the Certifying Examination, Pearson product-moment correlation coefficients were calculated for each medical condition in relation to each examination component. One hundred ten correlation coefficients were calculated, and of that number only ten reached significance at the .05 level of confidence. There were no significant correlations between

compliance with audit criteria and performance on the data gathering or management components of patient management problems. The only significant relationships observed were between compliance scores for some medical conditions and the two multiple-choice components. Correlations between vaginitis compliance scores and both multiple-choice components of the examination ranged from .46 - .52. For venereal disease the correlation between the compliance scores and the multiple choice management component equaled .36 and .39. Finally, the correlation between compliance scores for the cluster containing otitis media, sore throat, urinary tract infection and vaginitis and the two multiple-choice components equaled .38 and .46. All coefficients were significant at the .05 level of confidence.

However, in view of the large number of coefficients which were calculated, one should view the significance of these ten with some caution. It is generally accepted that in calculating a large number of correlation coefficients, some will be significant by chance alone. Although the presence of a significant correlation between a cluster of high-frequency primary care conditions and multiple choice scores provides encouraging evidence for the validity of the examination; the lack of correlation for other conditions such as headache and health maintenance makes interpretation of the data difficult. Thus, the conclusion most consistent with data presented in Table 17 is that, in general, the physician's assistants included in this study do not demonstrate a similar rank order in practice, as reflected by the medical record audit, as they did on the Certifying Examination.

A number of factors may have contributed to the lack of correlation between examination performance and compliance with audit criteria. While the medical content sampled on the Certifying Examination includes more than 50 different medical problems and conditions, the content of the medical record audit employed in this study sampled only twelve different medical conditions. It

Table 17

Correlation Coefficients Between Audit Scores
and Examination Scores

Audit Data	MCQ Interpretation	MCQ Management	PMP/Data Gathering	PMP Management
Angina				
Data gathering	-.09(12)	-.05	.45	--
Composite	-.03	-.03	.41	.32
Asthma				
Data gathering	-.35(15)	-.20	-.25	--
Composite	-.37	-.23	-.26	-.11
Birth Control Pills				
Data gathering	.87*(7)	-.44	-.03	--
Composite	.79*	-.46	.04	.62
Diabetes Mellitus				
Data gathering	-.40(9)	-.09	.33	--
Composite	-.40	-.04	.25	.59
Headache				
Data gathering	-.09(55)	-.03	.17	--
Composite	-.09	-.03	.17	-.07
H. Maintenance				
Data gathering	-.13(65)	-.12	.12	--
Composite	-.10	-.13	.14	.15
Hypertension				
Data gathering	-.08(33)	-.01	.05	--
Composite	-.09	-.08	.06	.00
Otitis Media				
Composite	-.04(83)	-.07	.18	-.07
Sore Throat				
Composite	-.12(102)	-.13	.04	.14

*indicates significance at the .05 level
( ) numbers in parentheses indicate the number of cases available for analysis
-- no correlation coefficients were calculated between data gathering and management data sets

Audit Data	MCQ Interpretation	MCQ Management	PMP/Data Gathering	PMP Management
Urinary Tract Infection				
Data gathering	-.18(65)	-.10	-.01	--
Composite	-.10	-.02	.04	-.04
Vaginitis				
Data gathering	.46(39)*	.48*	.27	--
Composite	.46	.52*	.27	.14
Venereal Disease				
Management	.06(36)	.39*	--	-.08
Composite	-.11	.36*	-.06	-.14
Diabetes Mellitus (Follow-up)				
Composite	.15(37)	-.14	-.13	-.13
Diastolic Hypertension (Follow-up)				
Composite	.03(76)	-.12	.13	-.06
Asthma (Follow-up)				
Composite	.01(20)	-.13	.07	.00
Birth Control Pills (Follow-up)				
Composite	.07(16)	-.18	-.31	-.48
Angina (Follow-up)				
Composite	.28(22)	-.34	-.30	-.23
Headache + Hypertension	-.07(21)	-.07	.21	-.03
H. Maintenance + Otitis Media + Sore Throat	-.10(49)	-.21	.24	.01
Otitis Media + Sore Throat + UTI + Vaginitis	.38(26)*	.46*	.05	.08

had been anticipated that this inability to control for the content in these two sources of data would make it difficult to evaluate the effectiveness of the Certifying Examination as a predictor of subsequent practice performance.

Previous evidence of the influence of medical content on performance had been obtained from the analyses of several certifying examinations. In an effort to determine whether medical content was also a factor in the extent of compliance with audit criteria, correlation coefficients were calculated among the compliance scores for the medical conditions included in this study. The resulting inter-correlation matrix is presented in Table 18.

One hundred eighteen correlation coefficients were calculated. Of that number, 77 (or 66%) did not reach significance at the .05 level of confidence. Of the 41 coefficients that did reach statistical significance, only 24 involved sample sizes of twenty or more cases. These coefficients ranged in magnitude from .31 to .78. Thus, there appears to be considerable variation in compliance with audit criteria across medical conditions. The fact that content appears to be a significant factor in determining performance both on examinations and on record audits poses significant problems for accurately assessing the true relationships between these two measures.

Another factor to be considered is that although medical record audits are coming to be widely used, there is not yet a consistent body of literature which supports the validity of this assessment technique. Although it seems unlikely that health professionals would systematically fabricate entries in the medical record, it seems reasonable to suspect that they do not document all pertinent behavior. Thus, what appears to be lack of compliance with audit criteria may simply be a failure to record all pertinent aspects of actual behavior. If such is the case, it raises a number of important questions concerning the validity of medical record audits.

	1 Angina	2 Asthma	3 Birth Control Pills	4 Diabetes	5 Headach	6 Health Mainten ance	7 Hyper- tension	8 Otitis Media	9 Tonsill pharyng	10 Urinary Tract Infecti	11 Vaginit	12 Venerea Diseas	13 Diabet (Follow-up	14 Hypert sion(Follc up)	15 Asthma (Follow-u	16 Birth control (Follow-up	17 Angina (Follow-up
1																	
2	—																
3	—	—															
4	*1.00 (3)	—	—														
5	* .94 (7)	* .82 (10)	.02 (7)	—													
6	* .65 (10)	.79 (5)	-.34 (6)	-.16 (3)	.32 (32)												
7	* .93 (5)	.87 (4)	-.37 (3)	* .97 (6)	.66 (21)	.52 (20)											
8	-.01 (8)	* .73 (12)	-.46 (7)	-.97 (3)	.58 (43)	.27 (50)	.51 (23)										
9	.39 (10)	* .68 (14)	-.03 (6)	.06 (5)	.50 (52)	.31 (61)	.18 (26)	.61 (82)									
10	.46 (7)	* .78 (10)	.01 (4)	-.71 (4)	.69 (38)	.23 (35)	.34 (22)	.78 (52)	.58 (62)								
11	.47 (4)	* .73 (8)	.07 (6)	—	.65 (27)	.41 (23)	.52 (14)	.51 (33)	.34 (37)	.64 (29)							
12	.93 (3)	.55 (8)	—	—	.62 (20)	* .60 (15)	* .70 (9)	.60 (23)	.71 (29)	.61 (20)	.44 (12)						
13	.64 (4)	—	.13 (4)	.13 (7)	.23 (18)	.32 (26)	.11 (12)	.38 (15)	.36 (31)	.39 (18)	* .72 (12)	.72 (4)					
14	.34 (9)	.98 (6)	-.21 (6)	-.14 (9)	.36 (34)	.18 (48)	.39 (28)	.42 (55)	.41 (65)	.54 (40)	.68 (22)	.41 (16)	.69 (35)				
15	.97 (3)	—	—	—	.72 (10)	.46 (13)	.32 (4)	.55 (17)	* .58 (19)	.46 (12)	* .92 (9)	*.98 (6)	.54 (11)	.46 (16)			
16	—	—	-.68 (4)	—	-.06 (11)	.28 (11)	-.16 (7)	.42 (13)	.16 (15)	-.13 (9)	.23 (11)	.52 (6)	.37 (9)	.44 (12)	-.05 (4)		
17	-.32 (4)	—	* .99 (3)	-.03 (5)	.01 (11)	.34 (16)	.34 (8)	.08 (14)	.31 (17)	.09 (9)	.00 (7)	—	.09 (18)	.37 (21)	.47 (10)	-.35 (6)	

— Indicates that less than 3 cases were available for analysis

( ) Indicates the number of cases used in calculating the correlation coefficient

* Indicates a correlation coefficient involving less than 20 cases, but significant at the .05 level of confidence

◯ Indicates a correlation coefficient involving more than 20 cases, and significant at the .05 level of confidence

Table 18

If, however, medical records can be assumed to accurately reflect the behavior of health professionals, then factors in addition to variability in content must be considered in attempting to explain the lack of correlation between examination performance and compliance with audit criteria.

One possible explanation is that what physician's assistants know about the diagnostic evaluation and management of patients and what they actually do in their respective practice environments may be significantly different. If their motivation in taking an examination is to demonstrate that they know what constitutes optimum care, their motivation in a practice setting may be to conform to patterns of health care delivery and record keeping which meet the expectations of the physician employer. If the resulting behavior is different in the practice setting from what it is in the testing situation, one cannot expect to observe significant positive correlations between the two.

If the differences in performance do exist between testing situations and practice settings and if medical record audits are valid assessments of behavior, then it may be unreasonable to expect that physician's assistants will fall in the same rank order on a medical audit as they do on a Certifying Examination.

Since a certifying examination seeks only to identify individuals who have not achieved minimum acceptable standards, then perhaps the same rationale should be applied to the assessment of compliance with audit criteria. If this viewpoint is accepted, then one would hope that individuals classified as acceptable or unacceptable in terms of their examination performance would be similarly classified as acceptable or unacceptable in terms of their compliance with minimum standards of care as reflected in a medical record audits.

The pass-fail level on the 1974 Certifying Examination was set at a standard score of 420 or sixty-four percent correct. For the purposes of

this comparison between pass-fail status on the examination and pass-fail status in relation to compliance with audit criteria, a sixty-four percent compliance rate was set for the record audit. A two-way contingency table was used to tally each physician assistant's pass-fail designation based upon his examination score and his audit compliance score. A chi square analysis was applied to these data as displayed in Table 19. The chi square value equaled zero, indicating that there was no relationship between pass-fail status on the examination and pass-fail status in terms of audit compliance.

As can be observed from Table 19, a rather large number of the participants (30) passed the examination but failed to meet the minimum audit compliance level. An inspection of the examination scores of this group of participants indicated that although many of them did achieve a passing score on the examination, the score was near the pass-fail level. For this reason, a new contingency table was prepared, as illustrated in Table 20, and the pass-fail level for the examination and audit compliance was adjusted upward from sixty-four percent to sixty-nine percent. The resulting chi square equaled 5.47 which was significant at the .02 level of confidence, thus indicating a significant relationship between pass-fail status on the Certifying Examination and pass-fail in relation to compliance with audit criteria.

Conclusions and Recommendations

On the surface, the results of this study would seem to indicate that the rank order of physician's assistants in terms of their compliance with audit criteria does not correlate significantly with their rank order in terms of scores on the Certifying Examination. However, in relation to minimum acceptable standards, a significant relationship appears to exist between the Certifying Examination and compliance with audit criteria when the minimum standard for both is set at 69 percent. Finally, there appears

Table 19

Contingency Table

Minimum Standards

EXAMINATION (64%)

AUDIT (64%)		Pass	Fail
	Pass	72	10
	Fail	30	4

$X^2 = 0.0$

Table 20

Contingency Table

Revised Minimum Standards

Examination (69%)*

AUDIT (69%)		Pass	Fail
	Pass	39	21
	Fail	26	30

$X^2 = 5.47$ ($p = .02$)

*If this standard had been applied to actual examination data, it would have resulted in a 50% failure rate.

to be considerable variation in compliance with audit criteria across the medical conditions included in this study.

However, these conclusions must be viewed as tentative in light of the fact that the validity of medical record audits has not yet been demonstrated nor have the environmental factors which impact practice behavior been identified. These issues are of concern not only for the development and validation of certifying examinations, but also for the assessment of continuing competence and recertification in the health professions.

If further studies show that medical records do not adequately reflect actual behavior, then other criterion measures will need to be developed in order to validate certifying examinations and assure continuing competence.

For these reasons, it is recommended that systematic studies be undertaken to investigate the validity of medical record audits and to identify factors which influence the record keeping behavior of health professionals. Once such studies have been completed, the predictive validity of the Physician's Assistant Certifying Examination should again be investigated with special emphasis on the control of content in the predictive and criterion measures.

1975 PHYSICIAN'S ASSISTANT CERTIFYING EXAMINATION

The third Certifying Examination was administered to 1411 candidates in 56 test centers on November 19-21, 1975. The content and format of the examination were equivalent to the 1974 examination program. The one-day written examination was divided into two major components: multiple-choice questions and patient management problems. In addition to the written portion of the examination, each candidate underwent assessment of his/her skill in performing selected components of a physical examination.

Items on the multiple-choice question section of the examination were designed to assess the candidate's knowledge and his/her skill in applying knowledge related to high priority health care functions. These items covered materials in the following broad areas of competence.

a. selection of laboratory and diagnostic procedures;

b. interpretation of clinical findings;

c. selection of management strategies;

d. patient counseling and instruction;

e. knowledge of medical sciences; and

f. knowledge of behavioral sciences.

Items on the patient management problems were classified into two subscales: data gathering and management/therapy. Candidates were presented with 13 simulated cases and asked to make decisions regarding the appropriate diagnostic workup and management of the patient as would be done in an actual clinical setting. These problems were designed to assess the candidate's skill in gathering pertinent information about patients and in making appropriate management decisions. The Certifying Examination has continued to focus on primary care, and for this reason the content of the examination was drawn from all of the clinical disciplines.

The assessment of physical examination skills continued to focus on each candidate's ability to examine the heart, lungs, eyes, abdomen, and neurologic system. As was the case in 1974, each candidate was evaluated by two physician examiners, each of which observed the candidate performing different components of the physical examination.

Table 21 summarizes the composition of the candidate group in 1975. There were no marked changes in the proportion of candidates from each of the four categories. Seventy-three percent of the candidates received their training in physician's assistant programs, eleven percent in Medex programs, and five percent in nurse practitioner programs. Eleven percent of the candidates qualified for the examination on the basis of work experience (informal training).

Table 21

Composition of Candidate Group

1975 Examination

Type of Training	Number	Percent
Physician's Assistant	1028	73
Medex	151	11
Nurse Practitioner	80	5
Informally Trained	152	11
Total	1411	100

For scoring purposes, four subsections of the examination were identified:

MCQ: represents the multiple-choice question portion of the examination.

PMP (D/G): represents those sections of the patient management problems that involved the selection of historical questions and physical examination procedures.

PMP (M/T): represents those sections of the patient mangement problems that involved the selection of laboratory studies and management decisions.

PAS: represents the physical examination skill assessment portion of the examination. (All examination components - heart, lungs, eyes, abdomen and neurologic - were weighted equally).

As shown in Table 22, the reliability of the total examination was .92, which places it within the range of reliabilities for other National Board examinations. The mean difficulty level for the multiple-choice (MCQ) section of the examination equaled .58 which indicates a slightly more difficult set

of questions compared to other National Board examinations. This statistic indicates that the average candidate from a formal training program answered about 58 percent of the multiple-choice questions correctly. The mean difficulty level of the patient management problems (PMP) was .79 for the Management Therapy section and .77 for the Data-Gathering section. These difficulty levels correspond to those found on other PMP examinations. On the Performance Assessment (PAS) portion of the examination, the average candidate from a formal training program obtained 75 percent of the total possible points. This represents an improvement in performance over that observed on last year's examination where the average candidate obtained 67 percent of the total possible points.

Table 22

Examination Statistics

1975 Examination

Statistic	
Composite Reliability	.92
Mean Difficulty	
MCQ	.58
PMP (D/G)	.77
PMP (M/T)	.79
PAS	.75

The statistical analysis of the 1975 examination indicated that the composite reliability and difficulty levels were comparable to those observed on the 1974 examination. The improved performance on the physical examination assessment as noted above was probably due to better candidate preparation for this portion of the Certifying Examination.

Table 23 summarizes the examination statistics that have been obtained during the past three years. The statistics were calculated on the reference group which, for the past two years, has been defined as candidates who are graduates of formal training programs and are taking the examination for the first time. The column labeled "average P" refers to the difficulty level of the examination and describes the proportion of items answered correctly by an average candidate. Reliability estimates have been calculated using the Kuder-Richardson 20 formula which is an estimate of internal consistency.

Table 23

1975 - Physician's Assistant Examination

Performance of the Examination

(Reference Group)

I. Item Analysis Data

A. Total Test

Year	Number of Items	Composite Reliability
1973	606	.89
1974	1266	.93
1975	1215	.92

B. MCQ

Year	Number of Items	Average P	Reliability
1973	147	.64	.83
1974	216	.57	.89
1975	210	.58	.87

C. PMP (Mgmt./Therapy)

Year	Number of Items	Average P	Reliability
1973	37	.82	.54
1974	284	.78	.76
1975	289	.79	.69

D. PMP (Data/Gath.)

Year	Number of Items	Average P	Reliability
1973	422	.79	.80
1974	554	.75	.89
1975	489	.77	.88

E. PAS (Performance Assessment)

Year	Number of Items	Average P	Reliability
1974	212	.67	.91
1975	227	.75	.93

Table 24 summarizes the intercorrelations among the various components of the 1975 Certifying Examination. These relationships are consistent with those observed on the two prior examinations, and suggest that each of the examination components is assessing aspects of competence that are different from those assessed by other components of the examination.

Table 24

Intercorrelation matrix of subscores

1975 Examination***

	PMP (Mgmt./Therapy)	PMP (Data/Gath.)	(Performance PAS Assessment)
MCQ	.57 (.73)	.37 (.42)	.27 (.29)
PMP (Mgmt./Therapy)		.32 (.41)	.17 (.21)
PMP (Data/Gath.)			.18 (.19)

***Correlations in parentheses have been corrected for unreliability.

Table 25 summarizes the performance by candidate group on the 1974 and 1975 Certifying Examinations. The composite score is derived by assigning equal weights to all components of the examination. Thus, each component contributes 25% to the total test score. It will be noted that candidates taking the examination for the first time score consistently and significantly higher than candidates who are repeating the examination. The informally trained candidates score consistently lower than candidates who are graduates of formal educational programs, and their performance on the 1975 examination declined from their level of performance in 1974. As might be expected, the failure rates for informally trained candidates and those who are repeating

the examination are considerably higher than failure rates for other groups of candidates. The nurse practitioners who registered for this examination continued to perform well, although they are too few in number to permit any generalization to the performance of nurse practitioners as a whole.

Performance on Physician's Assistant Certifying Examination

A. Performance on Total Test[1]

	Reference Group		First Takers		Repeaters		Informally Trained		Medex		P.A.		Nurses		Total Group	
	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.
1974	500	70	491	75	442	54	426	83	460	73	503	68	499	96	490	75
1975	500	70	489	78	384	7/	387	87	476	86	493	72	521	70	481	83

B. Performance on MCQ

	Reference Group		First Takers		Repeaters		Informally Trained		Medex		P.A.		Nurses		Total Group	
	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.
1974	500	100	490	105	390	78	414	110	419	97	510	95	488	116	487	105
1975	500	100	488	107	345	98	375	110	430	112	496	103	528	96	478	113

C. Performance on PMP - Mgmt./Trt.

	Reference Group		First Takers		Repeaters		Informally Trained		Medex		P.A.		Nurses		Total Group	
	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.
1974	500	101	496	103	448	89	461	114	476	118	501	96	530	98	494	103
1975	500	100	492	104	372	139	408	118	464	135	494	103	528	89	483	112

D. Performance on PMP - D/G

	Reference Group		First Takers		Repeaters		Informally Trained		Medex		P.A		Nurses		Total Group	
	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.
1974	500	100	494	103	467	79	450	112	485	102	499	100	502	105	493	102
1975	500	100	492	105	423	97	409	112	497	100	494	101	517	107	486	106

E. Performance on PAS Composite

	Reference Group		First Takers		Repeaters		Informally Trained		Medex		P.A.		Nurses		Total Group	
	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.	X	S.D.
1974[2]	500	100	486	112	464	100	379	126	460	103	504	100	472	151	485	112
1975[3]	500	100	487	111	398	137	358	142	512	104	492	102	514	104	480	116

[1]The composite score for the total test = [1 (PMP - MGMT./Ther.) + 1 (PMP - Data/Gath.) + 1 (MCQ) + 1 (PAS)] /4
[2]The PAS Composite for 1974 was weighted according to the length of each individual section.
[3]In 1975, each section of the PAS was standardized before the composite was computed. The PAS Composite was then restandardized to have a mean of 500 and a standard deviation of 100 after the 5 sections were added together.

The graduates of Medex programs performed as well as P.A. program graduates on the data gathering sections of patient management problems. In 1974, they scored significantly lower on the physical examination assessment component, but their performance was significantly better in 1975. In relation to the PMP management/treatment and multiple-choice question sections of the examination, the Medex graduates scored significantly lower than the graduates of P.A. programs. Similar score comparisons were also observed in 1974.

The examination standard set in 1974 and again in 1975 equaled a composite score of 420. At this composite score, the overall failure rate equaled twenty percent. However, as might be expected from the differences in examination performance, the failure rates varied across the several candidate groups. For example, in 1975, the failure rate for informally trained candidates equaled 60.5%, while the failure rate for Medex program graduates equaled 23.8% and for P.A. program graduates, 14.2%. The failure rate for the nurse practitioner group equaled 7.5%.

Beginning with the 1976 Certifying Examination, a certain proportion of the test will be drawn from the pool of previously used test items. Moreover, the number of test committees will be reduced from 4 to 2, one with responsibility for the patient management problems and the other with responsibility for the multiple-choice questions. Both committees will continue to be inter-disciplinary, including physicians, nurses, and physician's assistants. It is anticipated that these measures will not only streamline development of the examination itself, but enhance its cost effectiveness.

At the present time, the pool of multiple-choice questions totals 572 test items. They are distributed across clinical disciples in the following numbers:

	Items
Medicine	248
Pediatrics	121
Surgery	88
Obstetrics-Gynecology	36
Behavioral sciences	79

In addition to the multiple-choice questions, the pool contains 38 patient management problems distributed across the clinical disciplines as follows:

	Patient Management Problems
Medicine	14
Pediatrics	15
Surgery	2
Obstetrics-Gynecology	5
Behavioral sciences	2

APPENDICES

CASE D

Mrs. Arlene Burns is 34 years old. She is a new patient and has brought with her a summary of her medical history which was provided by her physician in another state. Her history is unremarkable except for the presence of recurring migraine headaches during the past 18 years.

In response to your inquiry concerning why she is coming to the office today, Mrs. Burns responds, "My migraines have been acting up lately and I'd like to get my prescriptions renewed. You'll find everything in my record. I've been taking Valium, codeine, and Cafergot."

TURN IMMEDIATELY TO PAGE 4.

SELECT ONLY ONE.

205. What's happening in your life that's stressing you?

205. Nothing in particular. As a matter of fact, we have just moved into a new house. The community is much better than the one we had been living in before moving 1500 miles to here. My husband's job couldn't be better in terms of money and position... that's why we came here in the first place. (TURN IMMEDIATELY TO PAGE 8)*

206. I'd be glad to renew your prescriptions, but first let's go over how often you need to take them.

206. Okay, what do you want to know? (TURN IMMEDIATELY TO PAGE 7)*

207. Your migraines are acting up?

207. Yes, for the last few months or so it seems like they have been getting more frequent than usual. (TURN IMMEDIATELY TO PAGE 8)*

208. I'd like to check you first before renewing your prescriptions.

208. Well, okay, if you think it's necessary (TURN IMMEDIATELY TO PAGE 5)*

SELECT ONLY ONE.

209. Have you had any dizziness or any symptoms other than your migraines?

209. No, just the same old pains. They just are occurring more often, that's all. (TURN IMMEDIATELY TO PAGE 9)•

210. Do your current headaches feel any different than your usual migraines?

210. No, just the same old pains. They just are occurring more often, that's all. (TURN IMMEDIATELY TO PAGE 9)•

211. How often have you needed to take your medication?

211. Lately, it seems like I am taking the Valium almost daily...about 4 times a day for the last few months. The other medicines, only when I need them, which recently had been for two headaches a week. Generally, I stick to not taking over 10 Cafergot pills a week. The last few weeks I've gone over a little. I've been out of the codeine for about two weeks, but I don't take it except rarely because I get sick to my stomach. (TURN IMMEDIATELY TO PAGE 10)•

212. Could you explain what you mean when you say your migraines are acting up?

212. Yes, for the last few months or so it seems like they have been getting more frequent than usual. (TURN IMMEDIATELY TO PAGE 9)•

SELECT ONLY ONE.

213. How often have you needed to take your medicine?

213. Lately, it seems like I am taking the [illegible] almost daily...about 4 times a day for the last few months. The other medicines, only when I need the which recently has been for two headaches a week. Generally, I stick to not taking over 10 Calergot pills a week. The last few weeks I've gone over a little. I've been out of the codeine for about two weeks, but I don't take it except rarely because I get sick to my stomach. (TURN IMMEDIATELY TO PAGE 10)*

214. What's happened in your life that's stressing you?

214. [illegible] in particular. As a matter of fact, we have just moved into a new house. The community is much better than the one we had been living in before moving here. My husband's job couldn't be better in terms of money and position ...that's why we came here in the first [illegible]. (TURN IMMEDIATELY TO PAGE 8)*

215. What is the usual pattern for your headaches and how has it been different lately?

215. [illegible] Otherwise, it's been the same. (TURN IMMEDIATELY TO PAGE 11)*

216. Are you taking your medicine correctly?

216. [illegible] taking them for almost 20 years. (TURN IMMEDIATELY TO PAGE 10)*

SELECT ONLY ONE.

217. Do you feel you are slipping up occasionally and, perhaps, going over the required dose?

217. I should know by now how to take my medications. God knows I've been taking them for almost 20 years! (TURN IMMEDIATELY TO PAGE 12)*

218. Are you taking one medication more heavily than the others?

218. No, not really. It's just that I'm getting the headaches more frequently. (TURN IMMEDIATELY TO PAGE 13)*

219. Do you find you need to increase your dose in order to get the same effect?

219. No, not really. It's just that I'm getting the headaches more frequently. (TURN IMMEDIATELY TO PAGE 13)*

220. How long have you needed to take your medication?

220. Lately, it seems that I am taking the Valium almost daily... about 4 times a day for the last few months. The other medicines, only when I need them, which recently has been for [illegible] headaches a week. Generally, I try to not taking over 10 Cafergot [illegible] a week. The last few weeks I've gone over a little. I've been out of the [illegible] for about two weeks, but I don't take it except [illegible] stomach. (TURN IMMEDIATELY TO PAGE [illegible])*

SELECT ONLY ONE.

221. Have you and your husband had satisfactory relationships lately?

221. All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep badgering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE 15)*

222. Well, it certainly doesn't sound like stress from what you describe, so it probably would be a good idea to go into more careful detail on your headaches. What is the usual pattern for your headaches and how has it differed recently?

222. Normally, I might have a headache every 3 months or so, but for the last six months, it's been increasing to the point that it's averaging one time per week. Otherwise, the headaches feel the same. (TURN IMMEDIATELY TO PAGE 11)*

223. It sounds like there've been a lot of changes in your life. That could certainly be unsettling.

223. That certainly is a possibility because lately I've been quite upset with everyone. (TURN IMMEDIATELY TO PAGE 14)*

224. Sounds like it could be a carry over from the work of moving. What do you think?

224. That certainly is a possibility because lately I've been quite upset with everyone. (TURN IMMEDIATELY TO PAGE 14)*

SELECT ONLY ONE.

225. Even though your headaches may feel the same to you, I think it would be a good idea to have a neurologist check you to make sure that it's nothing more serious. How does that sound to you?

225. Well, if that's what you think is best, set it up as quickly as possible, so I can get my medicine renewed. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

226. What is the usual pattern for your headaches and how has it been different recently?

226. Normally, I might have a headache every 3 months or so, but for the last six months it's been increasing to the point that it's averaging one [illegible] week. Otherwise, the headaches feel the same. (TURN IMMEDIATELY TO PAGE 11)*

227. Could you describe a typical headache?

227. All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep badgering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE 15)*

228. Can we go over what in your life may be distressing you?

228. Nothing in particular. As a matter of fact, we have just moved into a new house. The community is much better than the one we had been living in before moving here. My husband's job couldn't be better in terms of money and position... that's why we came here in the first place. (TURN IMMEDIATELY TO PAGE 3)*

SELECT ONLY ONE.

229. Sounds like something in your life may be bothering you.

229. Nothing in particular. As a matter of fact, we have just moved into a new house. The community is much better than the one we had been living in before moving here. My husband's job couldn't be better in terms of money and position ..that's why we came here in the first place. (TURN IMMEDIATELY TO PAGE 8)*

230. Taking that much Valium sounds like you may be getting yourself into other problems.

230. ...All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep hammering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE 15)*

231. Certainly does sound like your migraines are acting up! What is the usual pattern for your headaches and how has it differed recently?

231. Normally, I might have a headache every 3 months or so, but for the last 6 months it's been increasing to the point that it's averaging one time per week. Otherwise, the headaches feel the same. (TURN IMMEDIATELY TO PAGE 11)*

232. Why are you taking so much Valium?

232. All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep hammering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE 23)*

SELECT ONLY ONE.

233. What makes them better or worse?

233. [illegible] don't know. Other than the [illegible] occurring more often, I have no other symptoms I am aware of. [illegible] that helped when I was 18 years old still do. I've been taking the Valium [illegible] a day lately for the last few [illegible]. I try not to take more than [illegible] occasionally [illegible] the last few weeks I've [illegible] And the codeine I only [illegible] last resort because it usually makes me throw up. I've been out of them [illegible] the last two weeks. (TURN IMMEDIATELY TO PAGE [illegible])

234. I'd like to give you a physical examination to rule out serious causes for your headaches. If nothing turns up, I think we might want to start thinking about having you talk with Dr. Gray, our psychiatrist, to see if he might be able to help. Does that sound like a reasonable plan?

234. Well, if that's what you think is best. Set it up as quickly as possible so I can get my medicine renewed. (This is the end of the case. TURN IMMEDIATELY TO PAGE [illegible])

235. That's quite an increase in your symptoms. What do you think is making them more frequent?

235. [illegible] occurring more often. I have no other symptoms I am aware of. The [illegible] that helped when I was 18 years old still do. I've been taking the Valium [illegible] a day lately for the last few months. I try not to take more than [illegible] a week, but occasionally [illegible] the last few [illegible] slightly [illegible] the codeine I only take as a last resort because it usually makes [illegible] I've been out of [illegible] two weeks. (TURN IMMEDIATELY TO PAGE [illegible])

236. I'll do the physical examination now. If we don't find anything, we'll renew your prescriptions. The physical wil' tell us if anything more serious migh be going on.

236. [illegible] always and all [illegible] medication [illegible] have to do [illegible] [illegible] TO PAGE [illegible]

SELECT ONLY ONE

237. I'm sorry if I upset you. What I would like to determine is how often you need to take your medicine.

237. Lately, it seems like I am taking the Valium almost daily...about 4 times a day for the last few months. The other medicines, only when I need them, which recently has been for two headaches a week. [illegible] stick to [illegible] over [illegible] Cafergot pills a w[illegible] [illegible] gone over a little. I've been out of the codeine for about two weeks, but I don't take it except rarely because I get sick to my stomach. (TURN IMMEDIATELY TO PAGE 16).

238. You're angry with me for what I've said?

238. I guess I am a little. It seems like I've been more on edge lately with everyone. (TURN IMMEDIATELY TO PAGE 17).

239. I'm just checking. After all you and I are both human and can make errors.

239. All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep badgering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE 15).

240. You seem upset. Do you feel more on edge lately?

240. All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep badgering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE 15).

SELECT ONLY ONE.

241. You seem upset with me for asking that question.

241. Guess I am a little. It seems like I've been more on edge lately with everyone. (TURN IMMEDIATELY TO PAGE 17)*

242. What's happening in your life that's stressing you?

242. Nothing in particular. As a matter of fact, we have just moved into a new house. The community is much better than the one we had been living in before moving here. My husband's job couldn't be better in terms of money and position ... that's why we came here in the first place. (TURN IMMEDIATELY TO PAGE 9)*

243. What is the usual pattern for your headaches and how has it been different recently?

243. Normally, I might have a headache every 3 months or so, but for the last 2 months it's been increasing to the point that it's averaging one time per week. Otherwise, the headaches feel the same. (TURN IMMEDIATELY TO PAGE 11)*

244. Do you think you're taking your medicine correctly?

244. I should know by now how to take my medications. God knows I've been taking them for almost 20 years! (TURN IMMEDIATELY TO PAGE 12)*

SELECT ONLY ONE.

245. Valium may not be quite what you need. I'd like to give you something to help lift your mood too. Why don't I talk it over with the doctor, get you the prescriptions, and see you again when these headaches get better.

245. That's fine. Perhaps I can talk better when I get these headaches under control. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31).

246. Why don't we see if we can find out why you may be so sensitive lately?

246. All I came in here for was to get my prescriptions renewed for a condition I've had for 20 years and you keep badgering me with questions. Couldn't you just check my records and renew my medicine? (TURN IMMEDIATELY TO PAGE [illegible]5).

247. Moving certainly is a stressful period in one's life. Let me renew your medications and if you continue to have problems, please let me know.

247. That's fine. Perhaps I can talk better when I get these headaches under control. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31).

248. With everyone?

248. [illegible] that way, but I guess it's mostly with my husband and [illegible]. Probably it will go away after we are [illegible] settled [illegible]. [illegible] it's only been about 3 months since my husband, Jack, informed us [illegible] moving [illegible]. (TURN IMMEDIATELY TO PAGE [illegible]).

ERIC

SELECT ONLY ONE.

249. I apologize for angering you. I really do feel it would be safer if I went over your symptoms first in case anything new may have occurred. We can do this next week, if you like, after I renew your medications and things may have calmed down a bit.

249. That's fine. Perhaps I can talk better when I get these headaches under control. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

250. I'm afraid I couldn't do that until we have gone over your case. It could very well be something new and harmful to you. What has the pattern of your headaches been?

250. Normally, I might have a headache every 3 months or so, but for the last 3 months it's been increasing to the point that it's averaging one time per week. Otherwise, the headaches feel the same. (TURN IMMEDIATELY TO PAGE 11)*

251. I'm afraid I didn't realize how uncomfortable you currently are. Let me renew your medications and if you continue to have problems, let me know.

251. That's fine. Perhaps I can talk better when I get these headaches under control. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

252. I apologize for angering you. I really do feel it would be safer if I went over your symptoms first in case anything new may have occurred. What is the usual pattern for your headaches and how has it differed recently?

252. Normally, I might have a headache every 3 months or so, but for the last 3 months it's been increasing to the point that it's averaging one time per week. Otherwise, the headaches feel the same. (TURN IMMEDIATELY TO PAGE 11)*

SELECT ONLY ONE.

253. You've used up the codeine first. Would you like to talk about that?

253. I'm tired of having to go over everything plus your doubts too. I'm not and have no desire to be an addict. I've been hurting lately and upset with everyone. (TURN IMMEDIATELY TO PAGE 19)*

254. It certainly does sound like your migraines are acting up. Why do you feel you need more Valium?

254. I've been miserable from these headaches and really on edge with everyone. (TURN IMMEDIATELY TO PAGE 20)*

255. Frankly, I'm concerned about your increased use of Valium and codeine. Can we talk about that?

255. I'm tired of having to go over everything plus your doubts too. I'm not and have no desire to be an addict. I've been hurting lately and upset with everyone. (TURN IMMEDIATELY TO PAGE 19)*

256. Frankly, your headaches concern me and I would like to give you a complete examination now to see if I can discover anything. If nothing turns up, we can renew your medications and have you call back if you're not better in a few weeks.

256. Fine. I think they are the same as [illegible] all I want is the medication anyway, but if we have to go through this, I understand. (This is the end of the [illegible] TURN IMMEDIATELY TO PAGE 21)*

SELECT ONLY ONE.

257. What's happening in your life that's stressing you lately?

257. Nothing in particular. As a matter of fact, we have just moved into a new house. The community is much better than the one we had been living in before moving here. My husband's job couldn't be better in terms of money and position...that's why we came here in the first place. (TURN IMMEDIATELY TO PAGE 6)•

258. With everyone?

258. It seems that way, but I guess it's mostly with my husband and kids. Probably it will go away after we are completely settled from the move. After all, it's only been about 6 months since my husband, [illegible] informed us that we were moving. (TURN IMMEDIATELY TO PAGE 15)•

259. I'm sorry I upset you too. What I do need to determine is how often you need to take your medicines.

259. Lately, it seems like I am taking the Valium almost daily...about 3 times a day for the last few months. The other medicines, only when I need them, which recently has been for two headaches a week. Generally, I stick to not taking over 10 Cafergot pills a week. The last few weeks I've gone over a little. I've been out of the codeine for about two weeks, but I don't take it except rarely because I get sick to my stomach. (TURN IMMEDIATELY TO PAGE 16)•

260. I apologize for angering you. I really do feel it would be safer if I went over your symptoms in case anything new may have occurred. Perhaps for the time being I could renew your medications and we can talk next week when things have calmed down a bit.

260. That's fine. Perhaps I can talk better when I get these headaches under control. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)•

SELECT ONLY ONE.

261. How did you feel about it when he informed you?

261. He came in and proudly announced his fine promotion and salary raise of $7000 plus the move. It was pretty dramatic at the time. I guess I was pretty excited for him. (TURN IMMEDIATELY TO PAGE 21)*

262. Moving is certainly most stressful on the woman of the house. Valium may not be quite what you need to cope with what's going on. I'll prescribe something that's a little stronger. Check back with me in a couple of weeks when things have calmed down.

262. That's fine. Probably it will all blow over after I'm totally settled. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

263. Informed you that you were moving?

263. He came in and proudly announced his fine promotion and salary raise of $7000 plus the move. It was pretty dramatic at the time. I guess I was pretty excited for him. (TURN IMMEDIATELY TO PAGE 21)*

264. Moving certainly is a stressful period in one's life. I'll renew your medications and if you continue to have problems, let me know.

264. That's fine. Probably it will all blow over after I'm totally settled. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

SELECT ONLY ONE.

265. I'd like to give you a physical examination to rule out serious causes for your headaches. If nothing turns up, then I think we might want to start thinking about having you talk with Dr. Gray, our psychiatrist, to see if he might be able to help. Does that sound like a reasonable plan?

265. Well, if that's what you think is best. Set it up as quickly as possible so I can get my medicine renewed. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

266. I could feel comfortable renewing your Cafergot now, but I don't think I could give you Valium and codeine.

266. We don't seem to be communicating so well. I wonder if I could see someone else about my problem? (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

267. Upset with everyone?

267. It seems that way, but I guess it's mostly with my husband and kids. Probably it will go away after we are completely settled from the move. After all, it's only been about 3 months since my husband, Jack, informed us that we were moving. (TURN IMMEDIATELY TO PAGE 31)*

268. I'm sorry you're angry. Are you often this upset?

268. It seems that way, but I guess it's mostly with my husband and kids. Probably it will go away after we are completely settled from the move. After all, it's only been about 3 months since my husband, Jack, informed us that we were moving. (TURN IMMEDIATELY TO PAGE 31)*

SELECT ONLY ONE.

269. With everyone?

269. It seems that way, but I guess it's mostly with my husband and kids. Probably it will go away after we are completely settled from the move. After all, it's only been a short [illegible] since my husband, Jack, informed me that we were moving. (TURN IMMEDIATELY TO PAGE [illegible])*

270. Valium may not be quite what you need. I'd like to give you something to help lift your mood too. Why don't I talk it over with the doctor, get you the prescriptions, and see you again when we get these headaches under control.

270. Thanks. I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

271. I'm concerned about your increased use of Valium and codeine. Can we talk about that?

271. I'm tired of having to go over everything plus your doubts too. I'm not and have no desire to be an addict. I've been hurting lately and upset with everyone. (TURN IMMEDIATELY TO PAGE 18)*

272. I'd like to give you a physical examination to rule out serious causes for your headaches. If nothing turns up, we might want to start thinking about having you talk with Dr. Gray, our psychiatrist, to see if he might be able to help. Does that sound like a reasonable plan?

272. Well, if that's what you think is best. Just [illegible] as quickly as possible so I can get my medicine renewed. [illegible] end of the case. TURN IMMEDIATELY TO PAGE 31)*

SELECT ONLY ONE.

273. ...and then all the work of moving. I think I understand what you must be going through, so let me renew your medications.

273. That's fine. Probably it will all blow over after I'm totally settled. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

274. Well, it certainly does sound exciting. Has it continued to be that good for both of you?

274. Oh, yes. He's very happy in his work and in our new house which is much bigger than the old one. (TURN IMMEDIATELY TO PAGE 23)*

275. For him?

275. You know, it's really weird, but I never felt more depressed and angry, and I don't know why. As I said, he has a great new job and we have a new, bigger house. (TURN IMMEDIATELY TO PAGE 22)*

276. What about for you?

276. You know, it's really weird, but I've never felt more depressed [illegible] very sad, and I don't know why. As I said, he has a great [illegible] new [illegible] TO PAGE [illegible])*

SELECT ONLY ONE.

277. That anger certainly could be causing the headaches. Why don't I see if we can come up with something better than Valium to help you get control of things? First, I'd like to do a physical examination to make sure everything is okay.

277. That sounds fine. I certainly need something. (This is the end of the case. (TURN IMMEDIATELY TO PAGE 31)*

278. Some people have more trouble accepting happiness than pain in their lives. Perhaps this is occurring with you. If I find nothing wrong on physical examination, then why don't we think about having you work on this with Dr. Gray, our psychiatrist?

278. Well, if that's what you think is best. Set it up as quickly as possible so I can get my medicine renewed. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

279. Valium may not be quite what you need. I'd like to give you something to help lift your mood too. Why don't I talk it over with the doctor, get you the prescriptions, and see you again when we get these headaches under control?

279. That's fine. Perhaps I can talk better when I get these headaches under control. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

280. You feel angry and depressed over the move and don't know why?

280. Yes. I should be happy getting out of my hometown which is some 1500 miles from here. I've always thought of moving away from there [illegible] and now we have, plus having more money and I can't be but somehow it just wasn't right (TURN IMMEDIATELY TO PAGE...

SELECT ONLY ONE.

281. Some people have more trouble accepting happiness than pain in their lives. Perhaps this is occurring with you. If I find nothing wrong on your physical examination, then why don't we think about having you work on this with Dr. Gray, our psychiatrist?

281. I guess I'll have to think about it, but let's see what you find from the physical first. (This is the end of [illegible] TURN IMMEDIATELY TO PAGE 31)•

282. And in spite of all this good fortune, you're having headaches. I guess that makes me even more interested in having you see Dr. Brown, our neurologist, to make sure that something more serious is not going on.

282. Well, if that's what you think is best. Set it up as quickly as possible so I can get my medicine renewed. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)•

283. What about you?

283. You know, it's really weird, but I've [illegible] more depressed and [illegible] don't know why, as I [illegible] have a new, bigger house. (TURN [illegible])

284. Well, it sounds like it was a busy time, but now things should be settling down. So why don't I renew your medications after I check you and have you phone if there's any further problem?

284. [illegible] time. Probably it will all blow over after I'm totally [illegible] (This [illegible] the end of the case. TURN [illegible])•

SELECT ONLY ONE.

285. Maybe you just miss familiar places and activities.

285. Well...I was completing my dissertation in microbiology when this move came up and now I found out I can't complete it down here because the university doesn't offer that degree. (TURN IMMEDIATELY TO PAGE 25).

286. Wasn't right?

286. Well...I was completing my dissertation in microbiology when this move came up and now I found out I can't complete it down here because the university doesn't offer that degree. (TURN IMMEDIATELY TO PAGE 25).

287. Maybe you feel that way because you weren't involved in the decision to move.

287. Well...I was completing my dissertation in microbiology when this move came up and now I found out I can't complete it down here because the university doesn't offer that degree. (TURN IMMEDIATELY TO PAGE 25).

288. What seemed to be wrong?

288. Well...I was completing my dissertation in microbiology when this move came up and now I found out I can't complete it down here because the university doesn't offer that degree. (TURN IMMEDIATELY TO PAGE 25).

SELECT ONLY ONE.

289. How has that left you feeling?

289. I guess disappointed, hurt and angry. (TURN IMMEDIATELY TO PAGE 25)•

290. Aren't your children still pretty young and demanding most of your time?

290. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 27)•

291. Couldn't you try something else?

291. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 27)•

292. You must be really angry at your husband for making you move!

292. Not only my husband, but myself (TURN IMMEDIATELY TO PAGE 28)•

SELECT ONLY ONE.

293. What have you done to get over these feelings?

293. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 37)*

294. That's probably accounting for your headaches. Have you thought about ways to cope with your feelings?

294. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 37)*

295. Angry at your husband?

295. Not only my husband, but myself. (TURN IMMEDIATELY TO PAGE 37)*

296. Angry at whom?

296. Not only my husband, but myself. (TURN IMMEDIATELY TO PAGE 37)*

SELECT ONLY ONE.

297. That's a good idea. I'll finish your examination now and renew your medications.

297. Thanks. I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

298. I think you should discuss your feelings with your husband. I'll renew your prescriptions, and we'll make an appointment for next week to talk further.

298. Thanks. I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

299. Sounds like you're getting things under control. I'll finish up with the physical and then we'll get your medications renewed.

299. Thanks. I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

300. Perhaps it would be a good idea for you and your husband to come in and we could discuss this further. In the meantime I'll renew your prescriptions.

300. Thanks. I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31)*

SELECT ONLY ONE.

301. Angry at yourself? That must be an uncomfortable feeling.

301. Yes, it is. Maybe I shouldn't be so selfish, but I feel like I'm entitled to have my career too. I guess I didn't realize how important it was until now. (TURN IMMEDIATELY TO PAGE 29)*

302. I guess I could see why you're angry at yourself. You could try something else.

302. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 29)*

303. All that anger certainly can give you headaches. What do you think would make you feel less angry?

303. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 29)*

304. That doesn't sound so healthy for you. Have you thought of any solution?

304. I thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 29)*

SELECT ONLY ONE.

305. Perhaps when the children are grown you can return to your career.

305. I've thought about volunteering at the local hospital while my kids are in school. Maybe this will help take my mind off things and give me something to do. (TURN IMMEDIATELY TO PAGE 20.)

306. Perhaps you can try something else.

306. [illegible] about volunteering at the local hospital while my kids are [illegible]. Maybe this will help take [illegible] and give me [illegible]. (TURN IMMEDIATELY [illegible])

307. I can certainly see the bind. I'll finish the physical and then renew your medications to help you cope.

307. [illegible] appreciate your help. [illegible] (TURN IMMEDIATELY TO PAGE 21.)

308. Do you and your husband talk over major decisions like a move?

308. [illegible]ally. Although I care for [illegible] (TURN IMMEDIATELY TO PAGE 20.)

SELECT ONLY ONE.

309. We will have to be ending for now. In the meantime I wonder if you might not try being more assertive at home.

309. Thanks. I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31).

310. Perhaps if you and your husband come in, we could work on this together.

310. [illegible] I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31).

311. I'd like to talk with you more next week. In the meantime you might think about numerous alternatives, including separation.

311. [illegible] I appreciate your help. (This is the end of the case. TURN [illegible] TO PAGE 31).

312. Why don't you talk this over with your husband and let me know how things work out by phone next week?

312. [illegible] I appreciate your help. (This is the end of the case. TURN IMMEDIATELY TO PAGE 31).

APPENDIX B

PHYSICIAN'S ASSISTANT IPS OBSERVATIONAL FORM

Questioning	1. Non-case pertinent, open ended. 2. Case pertinent, open ended. 3. Non-case pertinent, closed (set of possible answers is predictable). 4. Case pertinent, closed (set of possible answers is predictable). 5. Case pertinent, inappropriate leading question. 6. Reflects patient comment or question as a probe. 7. Repetitious question. 8. Continuous questioning.
Reactions to Patient Behaviors	9. Inappropriately interrupts patient/abrupt shift in focus. 10. Encourages patient to continue talking. 11. Uses technical terminology. 12. Follows-up pertinent cues (new information). 13. Fails to follow-up pertinent cues (new information). 14. Clarifies own prior question/comment (e.g., "what I really mean is"). 15. Repeats own question verbatim. 16. Appropriately summarizes/synthesizes. 17. Inappropriately summarizes/synthesizes. 18. Expresses understanding. 19. Responds with inappropriate affect (e.g., disinterest, judgment, etc.).
Provision of Support/ Information	20. Reassures patient appropriately. 21. Reassures patient inappropriately. 22. Gives advice/instructions appropriately. 23. Gives advice/instructions inappropriately. 24. Answers patient's questions directly. 25. Answers patient's questions evasively. 26. Provides appropriate information. 27. Provides inappropriate information. 28. Expresses empathy.

SEQUENCE OF ACTIVITIES

I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV

APPENDIX C

AUDIOVISUAL SIMULATION STUDY

Patient Protocol Form

The following categories are presented in order to assist you in developing a patient protocol for the pilot studies in interpersonal skills. The information you provide will need to be sufficiently detailed so that an actor or actress can develop the appropriate characterization. Particular attention should be given to data which are of special relevance to the evaluation of this patient by the P.A.

These categories are meant to provide guidelines only. You may add other categories or omit any outlined here as you think appropriate. In developing a protocol for each patient, please organize the data under headings of some kind. This will assist us in identifying any additional work that may be needed on the protocol, and will also aid the actor in his assimulation of the role.

I. PERSONAL INFORMATION

Name
Sex
Age
Birthdate
Birthplace
Marital Status
Number of Children (their ages and sex)
Religion
Occupation
Spouse's name, age, occupation

II. CLINICAL SETTING

Initial patient visit or follow-up?

Purpose of interaction (primarily history taking?, counseling?, explaining?,)
Patient problems (those that precipitated the present need for health care)

III. PHYSICAL & EMOTIONAL HISTORY

Medical history
Drug history
Psycho-social history

IV. CURRENT PHYSICAL & EMOTIONAL STATUS

V. PAST & CURRENT LIFE SITUATIONS

Childhood family life & experience
Adult family life & experience
Current life situation and relationships

VI. AFFECTIVE PROFILE

VII. Appropriate P.A. behavior in response to patient's affective profile acceptance, understanding, reassurance

VIII. Type of information that should be provided to patient

IX. Type of Information that should be elicited from the patient

NATIONAL BOARD OF MEDICAL EXAMINERS

TONSILLOPHARYNGITIS, TONSILLITIS OR PHARYNGITIS

Records to be Abstracted: The following criteria are appropriate for the initial work-up of a patient with this condition.

HISTORY

1. Symptoms:

   Acceptable:
   - Sore throat, swollen throat
   - Painful or difficult swallowing
   - Headache
   - Coated or "strawberry" tongue
   - Anorexia
   - No associated symptoms

2. Duration of Symptoms:

   Acceptable:
   - Sore throat x _____ days
   - Malaise x 24 hr

3. Mononucleosis (if under 30 years of age):

4. Current Medications:

   Acceptable:
   - ASA
   - Antibiotic
   - No Rx

5. Previous Drug Reactions if Antibiotics are Prescribed: No manifestations are necessary.

   Acceptable:
   - None
   - NKA
   - Allergic to:_____
   - Penicillin allergy

PHYSICAL EXAMINATION

6. Throat Findings: Evidence that throat was examined.

   Acceptable:
   - Throat inflamed, injected
   - Tonsils inflamed, injected
   - Exudative findings
   - Tonsillitis, etc. under physical findings
   - Throat normal
   - Throat red

7. Temperature Recorded:

Acceptable: (Listed in degrees)

Unacceptable: Patient febrile

8. Ears: For patients under 25 years of age. NOT APPLICABLE if patient is 25 yrs. old or older.

Acceptable: Ears normal
TM's normal
TM's not inflamed, injected
Rest of ENT normal

9. Lymph Nodes: For patients under 25 years of age. NOT APPLICABLE if patient is 25 yrs. old or older. A global description or any reference to head and neck (cervical and/or submandibular nodes). Mention of the nodes outside the head and neck is insufficient by itself.

Acceptable: Anterior or posterior cervical tender, swollen, prominent Adenopathy

MANAGEMENT PLAN

10. Correctness of Action Concerning Drug: The correctness of the action taken is determined by the pattern of response to item 6 above and the ordering of a throat culture. Appropriate patterns are presented in the table below. Providing the patient with a temporary prescription, for 2 days or less, pending the return of the throat culture is an acceptable behavior. A refillable prescription or a notation that the patient will be told to stop the medication if the TC results are negative is also acceptable.

Patterns of responses not listed in the following table or described in the above paragraph are considered inappropriate and the "NO" column for item 10 should be checked.

Throat Findings	TC ordered	TC results	Indicated Therapy	Duration
+	Yes	+	Pc	10 days
+	Yes	-	None	NA
+	No		Pc	10 days
-	No		None	NA
-	Yes	+	Pc	10 days

If the duration does not appear in the record, and if no statement regarding future patient contact upon return of the culture is present, assume that the drug has been given for 10 days and check the "NO" column for item 10.

11. Drug: Acceptable drugs include penicillin - G or V, erythromycin, bicillin. Other drugs are acceptable if justified by allergies, additional organisms cultured, or Pc failure. NOT APPLICABLE if no drug prescribed.

12. Dosage: NOT APPLICABLE if no drug prescribed.

13. Duration: NOT APPLICABLE if no drug prescribed. Drug must be prescribed for 10 days if no culture was taken or if the results are positive. If the culture is not back, the prescription should be for 2 days or less or some notation regarding contact with the patient upon return of the culture should be made.

The following examples illustrate the appropriate coding of items 10-13.

Example 1.

Throat red. TC to lab

Rx Benylin Dinetapp

	Y	N	UC	N/A	MD
10.	☑	☐	☐	☐	☐
11.	☐	☐	☐	☑	☐
12.	☐	☐	☐	☑	☐
13.	☐	☐	☐	☑	☐

Example 2.

Throat is hyperemic. TC taken

Given Pen. G 400 for 7 days.

	Y	N	UC	N/A	MD
10.	☐	☑	☐	☐	☐
11.	☑	☐	☐	☐	☐
12.	☑	☐	☐	☐	☐
13.	☐	☑	☐	☐	☐

Example 3.

1/2/75 Tonsil hypertrophic & hyperemic. Rx: V-cillin 500 & TC done.

1/6/75 TC neg.

	Y	N	UC	N/A	MD
10.	☐	☑	☐	☐	☐
11.	☑	☐	☐	☐	☐
12.	☑	☐	☐	☐	☐
13.	☐	☑	☐	☐	☐

REFERENCES

Harrison's Principles of Internal Medicine, 6th Ed., Ch. 140, pp. 785-789.

Henry L. Barnett. Pediatrics, 15 Ed., Ch. 14.17, pp. 647-652.

APPENDIX D

NATIONAL BOARD OF MEDICAL EXAMINERS

ANGINA

Records to be Abstracted: The following criteria are applicable to the initial work-up of patients with angina.

Chart Selection or Diagnosis: The following list of sign out diagnoses is meant to be suggestive only and not inclusive. Sign out diagnoses may vary considerably and still be acceptable for this study. A chart must show either chest pain or angina to be acceptable for the study. If the diagnoses listed as numbers 3-7 below are present on a chart, they are acceptable as long as they are (or have been) accompanied by a notation regarding chest pain and/or angina.

1. History of exertional chest pain
2. Angina Pectoris
3. Coronary Artery Disease
4. Ischemic Heart Disease
5. Coronary Arteriosclerosis
6. Aortic Stenosis with Angina
7. Coronary Insufficiency

HISTORY

1. Character (quality) of Pain:

   Acceptable: Any qualitative description such as pressing, squeezing, etc.
   Dull, heavy
   Constricting
   Pressure
   Tearing

   Unacceptable: "Anginal pain"

2. Location or Radiation:

   Acceptable: Substernal (any anatomic description)
   Anterior chest
   Retrosternal ($\bar{c}$ or $\bar{s}$ radiation)
   Radiates to neck, jaw
   left
   both arms
   into back

P.A. Val.
3/2/76

3. Duration of Pain:

Acceptable:
30 seconds to a minute
1-2 minutes
About 3 minutes

4. Precipitating Factors, Mode of Onset:

Acceptable:
Brought on by 1 flight of stairs
On effort, during tense periods
Brought on by emotion only
After meals
While lying down
With intercourse
During day at work
Whenever I exert myself
Gradual, with exertion
Sudden, with increasing intensity

5. Associated Cardiac Symptoms: Presence or absence of symptoms.

Acceptable:
Accompanied by shortness of breath (S.O.B.)
Denies associated symptoms
PND
Diaphoresis
Weakness
Palpitations
Anxiety
Orthopnea
Nausea and vomiting

6. Time of Onset or Duration of Illness:

Acceptable:
When patient was 30 years old
9/15/71
Never before
Increasing frequency in past 6 months
10 years
6 months ago
First time yesterday evening

7. Frequency of Angina:

Acceptable:
Increasing frequency in past 6 months
2-3 times per week
2-3 times per day and at night

P.A. Val.
3/2/76

8. Personal History of Smoking: Specific Quantitation

Acceptable: 1 ppd
Non-smoker

Unacceptable: Smoker (insufficient by itself)

9. Personal History of Elevated B.P.:

Acceptable: Denies elevated B.P.

10. Personal History of Diabetes:

Acceptable: Denies diabetes
Diabetic x 4 yr.

11. Personal History of Elevated Lipids:

Acceptable: Lipids or cholesterol

12. Family History of Heart Disease:

Acceptable: Family hx negative

13. Activity Level of Patient:

Acceptable: Sedentary office worker
Construction worker
Jogs daily

14. Current Medications:

Acceptable: None
List of medications

15. History of Obesity if the Patient is Obese: See general description of patient in P.E. to determine if patient is obese, if some mention is not made in Hx. NOT APPLICABLE if patient is not obese.

Acceptable: Overweight all of life
Gained 50 pounds in last 6 months
(Recent weight change)

16. Emotional History: Any reference to emotional status or life stresses such as work, death, divorce found in the review of systems (ROS), social history (SH), or history of present illness (HPI) is acceptable.

Acceptable: Anxious
Handles stress well
Has occasional depression
Recent death in family

PHYSICAL EXAMINATION

17. Sex and Weight: Weight in pounds.

Acceptable: 160 lb. male

18. Nutritional Status:

Acceptable:
Obese
Well built
Thin
WN

19. Blood Pressure:

Acceptable:
140/86
220/136
220/120/60

20. Pulse Rate:

Acceptable:
76/min
130/min
76

21. Fundi:

Acceptable:
Normal
Grade 1 + A.S. (arteriosclerosis)
Arteriolar narrowing bilaterally
Marked increased arteriolar reflex

22. Cardiac Size:

Acceptable:
Not enlarged
Description of PMI (point of maximal impulse)location

Unacceptable: Heart normal

23. Cardiac Rhythm:

Acceptable:
Regular
Normal Sinus Rhythm (NSR)
Irregular

Unacceptable: Heart normal

24. Cardiac Auscultation: Sounds - normal or describe quality.

Acceptable: No (N)
Gallops - description
Description of (N)
No rubs, gallops
1st and 2nd heart sounds normal
Normal heart sounds

Unacceptable: Heart normal

25. Lungs - only auscultation required:

Acceptable: Clear to P&A
Bilateral basilar rales
Lungs normal
Chest clear

26. Carotid Pulses:

Acceptable: +2
Normal

27. Peripheral pulses - quality of pulses:

Acceptable: Presence or absence of bruits, lags
Pulses intact
(Diagram of pulses)
Bounding pulses present
Diminished pulses

LABORATORY AND DIAGNOSTIC STUDIES

If one or more of the following studies is found under Plan, it is acceptable and the appropriate criterion should be checked "YES".

28. ECG: Description of reading - an ECG alone w/o description is unacceptable.

Acceptable: Evidence of old M.I.
ST segment depression
ST segment elevation

29. Hemoglobin or Hematocrit:

Acceptable: Normal
Hematocrit 43%, Hemoglobin 14 g/ml

30. Blood Glucose (sugar):

Acceptable: Normal
SMA-12

31. Lipids: Cholesterol, Lipid Profile, Triglycerides - Any one of these is acceptable.

Acceptable: Lipids elevated for age
Elevated triglycerides

32. Chest X-ray: PA view and interpretation to R/O rib fracture, embolus, pneumonia infiltrate, tumor, pneumothorax, CHF.

Acceptable: Normal heart size

MANAGEMENT PLAN

33. Dietary Instructions if Obese, Hypertensive, or Diabetic: NOT APPLICABLE if patient is not in one of these three categories.

Acceptable: Patient instructed on diet
Referred to dietician
Specific low calorie diet

34. Patient Education: Instruction and counseling regarding such things as smoking, risk factors, and/or exercise.

35. Medications: A listing of medications is acceptable.

Acceptable: PRN Nitroglycerine
TG gr 1/150 PRN
Isordil 10 mg q 6 hr
Inderal 20 mg qid

36. Follow-Up Appointment: A note that an appointment has been made and the date (e.g., 2 weeks, 7 months) is required.

Acceptable: RTC 1 month
Re ✓ 1 month

Unacceptable: Return PRN
(No reference to return visit)

REFERENCES

Diagnosing Chest Pain: A panel discussion in Patient Care, Nov. 1970.

Harrison's Principles of Internal Medicine .

Oglesby, P. Severe Angina Pectoris. Hospital Medicine, March 1969.

Riseman, J.E.F. Clinical Course of Angina Pectoris. American Journal of Medical Science, 1966, 252, p. 146.

APPENDIX E

NATIONAL BOARD OF MEDICAL EXAMINERS

VALIDATION STUDY OF THE 1975 PHYSICIAN'S ASSISTANTS CERTIFYING EXAMINATION

INSTRUCTION MANUAL FOR RECORD REVIEWERS

Introduction

The purpose of national examinations for licensure and certification is to identify those candidates who have not yet achieved minimum acceptable levels of knowledge and skill required to deliver adequate health care. If examinations of this kind are to achieve their purpose, the content and format should be relevant to those functions the health professional will perform in practice. Moreover, there should be some positive relationship or correlation between a candidate's performance on the certifying or licensing examination and his actual performance in practice.

To date, however, national examinations in medicine and the allied medical professions have not been validated by comparing examination performance with subsequent performance in practice. Validation studies of this kind have not been conducted primarily because feasible evaluation methodologies for accurately and reliably assessing day-to-day performance have been lacking. Recent experience with medical record audits suggests that they may be a useful tool for this purpose. The use of the medical audit is based on the assumption that the entries made in the medical record reflect the care provided to the patient and/or the patient's family by physicians and other health care personnel. This assumption is consistent with the long standing purpose of the patient's medical record, namely to document the care provided.

While it appears that a review of medical records can be used as a methodology for evaluating some aspects of day-to-day performance, it should be recognized that there is a body of knowledge, skills, and attitudes relevant to the competence of health professionals that the medical audit cannot assess. For example, it is difficult to determine from the medical record the degree of the physician's empathy for the patient, the accuracy of the physician's skill in palpating an abdomen, or the patient's willingness to comply with various treatment regimens. These aspects of physician functioning are difficult or impossible to assess with the usual paper and pencil test as well. For this reason, it should be anticipated that the physician assistant's performance as measured by the medical record review will correlate with only some portions of the Certifying Examination and that neither methodology will measure all relevant aspects of performance.

Purpose and Method

The overall purpose of this validation study is to determine the extent to which performance on the Physician's Assistant Certifying Examination is predictive of an individual's subsequent performance in actual practice as measured by a medical record review.

NATIONAL BOARD OF MEDICAL EXAMINERS

In developing the protocol for this study, the NBME project staff asked a group of physicians and physician's assistants to select a number of clinical problems which physician's assistants frequently encounter and then to specify the information necessary for the diagnosis and management of patients with each of the clinical problems.

Responsibilities of the Reviewer

Retrieving and Coding Medical Records:

The reviewer is responsible for obtaining the records to be abstracted from a member of the clinic staff and completing a coding form for each record identified by staff personnel as appropriate for this study.

In abstracting data from a chart, you must determine whether data have been entered in a patient's medical record which satisfy the previously established criteria. For each criterion, you must check one of the following options on a coding form: YES if the medical record entry indicates that the criterion has been met; NO if the criterion has not been met; UNCERTAIN if you are uncertain if the criterion has been met (because of problems of legibility or interpretation of criteria); and NOT APPLICABLE if the criterion is not applicable for the situation at hand.

A fifth column, MD, has been added. If you can distinguish between PA and MD entries and note that a specific piece of information has been entered by the physician, check the column labeled "MD" in addition to the YES column. If you leave the "MD" column blank, it means the entry was made by the PA. If you cannot tell who entered the information in the record, check the box appearing at the end of the coding form labeled "I could not distinguish between MD and PA entries on this record".

Confidentiality:

Rules regarding the handling of Medical Records are of the utmost importance. The following apply to this study:

1. Never open a patient's Medical Record unless it is a patient on your list. When a record is reviewed, do so in a manner such that the record cannot be seen by patients or office employees.

2. Never discuss the contents of a Medical Record.

3. Never leave a Medical Record unattended where it may be seen by patients or others.

4. Treat all information contained in a Medical Record in a strictly confidential manner.

National Board of Medical Examiners

Do's and Don'ts for Reviewers:

1. Where the initial work-up for a diagnosis, such as hypertension or diabetes mellitus, was completed in the hospital, use the appropriate set of follow-up criteria in relation to follow-up evaluations of the patient in ambulatory care settings.

2. Sets of criteria for intial work-ups should be applied only to patients seen in the office or clinic for their initial evaluation.

3. Try not to use casual drop-in visits. An example of such a visit would be the visit of a patient who is just passing through the community of the practice. It seems rather unlikely that such a patient will return to the practice and, therefore, the criteria developed for this study are inappropriate.

4. If patients are new to a practice, the selection of the appropriate set of criteria depends on the availability of information collected previously by another practitianer. If there is a notation in the record for the initial visit that prior medical information is to be obtained, apply the appropriate set of follow-up criteria to the initial visit. If prior medical information has already been obtained, use the set of criteria for the initial visit and accept all data in the record that satisfy the criteria. If there is no indication that the practitianer will seek prior information, the set of criteria for the initial visit is to be used.

5. Study the criteria carefully immediately before beginning to review any medical records in a given medical condition.

6. Review all charts for a given medical condition before going on to another medical condition. For example, review the criteria and instructions for hypertension. Collect all patient records that are to be reviewed for hypertension and review all records before beginning to review the criteria for another medical condition such as health maintenance for adults.

7. Use pencil to code, to facilitate corrections you may wish to make.

8. Mark only one category (YES, NO, UNCERTAIN, or NOT APPLICABLE) for each criterion title.

9. If you can determine (usually by handwriting) that the entry was made by the physician, mark the MD category in addition to the YES or UNCERTAIN column.

10. If age is missing from a medical record, mark the UNCERTAIN column for all age dependent criteria.

11. If information is present in the chart but the information is not applicable for a patient of that age, mark the NOT APPLICABLE column. In other words, NOT APPLICABLE takes precedence over actual chart entries.

NATIONAL BOARD OF MEDICAL EXAMINERS

Reliability of Abstracts:

In order to make equitable judgments about performance on the basis of the medical record audit, there must be consistency in the manner in which the record entries are judged and entered on the coding forms. Thus, if several reviewers code the same records, their coding form should be consistent with one another. If a reviewer reabstracts the same set of record entries on two separate occasions, the reviewer's entries on the second coding form must match the entries made on the first record review. In addition, reviewers working at one site should interpret the criteria in the same way as reviewers working in another region.

This manual has been prepared to help you make reliable judgments. On the next few pages you will find some rules and procedures that apply to all problem areas. Following the general rules is a section specific to each clinical problem. Each set of problem specific criteria includes a list of criteria (e.g., Duration of Present Illness, Symptom of Sore Throat. etc.) and criteria instructions. The criteria instructions provide you with guidelines about how the criteria should be interpreted, examples of acceptable and unacceptable entries, and notations concerning the applicability of the criterion. The examples of acceptable and unacceptable entries are not intended to be exhaustive lists of chart entries. A P.A. may use any of a large number of synonyms to express his/her impressions. You must refer frequently to the instructions when abstracting records. The criteria items alone are not sufficient to ensure the reliability of the judgments being made.

Obtaining and Submitting Coding Forms:

Mail all completed coding forms to:

Mrs. Arleen Caba
National Board of Medical Examiners
3930 Chestnut Street
Philadelphia, Pennsylvania 19104

All requests for blank coding forms should also be directed to Mrs. A. Caba. Please make these requests in sufficient time to allow for duplicating and mailing.

Please submit all completed forms immediately after completing them. Include your Time Reporting Form and the office report form with your completed coding forms. Envelopes are available from Mrs. Caba.

Problems and Questions:

Any questions pertaining to the criteria and the medical records should be referred first to Dr. Vivian Erviti at the NBME. If Dr. Erviti is not available, questions should be referred to Ms. Jane Bunce. If neither Dr. Erviti or Ms. Bunce is available, please leave your message with Mrs. Arleen Caba. Dr. Erviti or Ms. Bunce will return your call as soon as possible.

Vivian Erviti, Ph.D.
National Board of Medical Examiners
3930 Chestnut Street
Philadelphia, Pa. 19104
(215)349-6400, Ext. 271

Ms. Jane Bunce
National Board of Medical Examiners
3930 Chestnut Street
Philadelphia, Pa. 19104
(215)349-6400, Ext. 269

12. Review the coding form to ensure that all criteria are completed and all identifying information has been entered. In completing the top of the coding forms, be sure all data are right adjusted, i.e., are placed in the blank spaces farthest to the right first. If the site you are visiting has been assigned #15, place this number on the lines below which "4" and "5" appear. If a log # or patient I.D. is 1049, place these figures on the lines 11, 12, 13, 14. Leave the lines numbered 6-10 blank. Dates should be entered as _9_ _2_ 7 5 for September 2, 1975. Enter your assigned number
15 18 20
in spaces 34-36. When a coding form has more than 1 page, be sure to complete the top of the second and succeeding pages.

13. If an item is NOT APPLICABLE because of age or sex, mark N/A even when an entry for the item exists.

14. Explain why you have marked the UNCERTAIN category for any criterion title (a note beside the title, at the top or end of the coding form will be satisfactory). This will aid the NBME staff in identifying problems with the criteria.

15. Give an estimate of the time it took you to abstract a record at the end of the coding form.

16. Prepare a separate listing for records pulled that cannot be used for abstracting. For each non-abstractable record pulled, enter the problem area and the reason for not abstracting the record.

17. Place all additional written comments at end of coding form or on a separate sheet of paper. These comments will aid us in the revision of criteria and the subsequent preparation of manuals for record reviewers.

Calculating the Age of Patients:

Some sets of criteria include several titles that are age-specific. Therefore, at times it will be necessary to obtain the age of the patient in order to determine whether or not a title is applicable to the record being reviewed. Age to the nearest year is sufficient.

If the age is missing from a chart being audited and there are age-specific criteria, audit the chart but mark each age-specific criterion UNCERTAIN. Note on the top of the coding form that the age was unavailable. In certain cases where age is missing, another entry in the record may make the general age range obvious. In these instances age-specific criteria can be checked either YES or NO. Please note on the top of the coding form the phrase used to determine the age range. An example in this latter category might be a notation that the patient is retired. Obviously, such a patient is above a 35 or 40 year old minimum.

Interpreting the age of Patients:

The interpretation of NOT APPLICABLE under 25 years of age will be that the criterion should be satisfied for a patient exactly 25 years old and older, but

would not be applicable for patients even 1 day less than 25 years old. The interpretation of NOT APPLICABLE after 25 years of age will be that the criterion will not apply for patients exactly 25 years old and older.

Interpreting the Criteria:

Entries found under the Review of Systems (ROS) are acceptable for history items.

A statement indicating either the presence or absence of any of the factors listed as examples of appropriate chart entries is acceptable because statements in either direction are valuable. Thus "no fever", for example, should be checked YES.

As stated previously, the lists of acceptable entries following each item in the criteria sets represent a few of the many ways in which signs and symptoms can be combined with other descriptive phrases. The lists are not all inclusive and the abstractor should check an item "YES" if any one of the key phrases appears, regardless of the presence or absence of additional detail that might accompany it. Please refer frequently to the instructions when abstracting records.

INTERPRETATION OF LABORATORY DATA*

Simultaneous Multispecimen Analysis (SMA)

SMA-6	SMA-12	SMA-C-20
Sodium	Total Protein	(all entries for SMA-12 plus the following:)
Potassium	Albumin	
Chloride	Globulin	
BUN	Inorganic Phosphatase	SGPT
Glucose	Bilirubin	CPK
	SGOT	Iron
	LDH	Triglycerides
	Alkaline Phosphatase	Total Protein
	Calcium	
	Glucose	
	Cholesterol	
	BUN/Cretinine	
	Uric acid	

(NOTE: Urinalysis (U/A) -- grams percent (gm %) or 4 +)

* As reported by The Children's Hospital of Philadelphia

NATIONAL BOARD OF MEDICAL EXAMINERS

P.A. VALIDATION STUDY

ANGINA FOLLOW-UP VISITS 1 7 / 1 2

P.A.# _____ 3 4 5

Pt. Sex _ Male (1) / Female (2) 27

Log # or Pt. I.D. 6 _____ 14

Date of Abstracting Mo Day Yr _29 _31 _33

Date of Patient Visit Mo Day Yr _16 _18 _20

Abstractor _____ 34 35 36

Age of Patient _ _ years 21 22 23

No. of Abstract _ First (1) / Recheck (2) 37

		Y	N	UC	N/A	MD
	HISTORY					
1.	Complaints:	☐	☐	☐	☐	☐ 38
	PHYSICAL EXAMINATION					
2.	Blood Pressure:	☐	☐	☐	☐	☐
3.	Pulse Rate:	☐	☐	☐	☐	☐
4.	Cardiac Auscultation:	☐	☐	☐	☐	☐
5.	Cardiac Rhythm:	☐	☐	☐	☐	☐
6.	Chest Auscultation:	☐	☐	☐	☐	☐
	MANAGEMENT PLAN					
7.	Continued Management Plan:	☐	☐	☐	☐	☐

☐ I could not distinguish between MD and PA entries on this record. 73

Time required _ _ _ mins. 74 – 76 P 77 A 78 3 79 1 80

OE 0068 423

REFERENCES

1. Bales, R.F.: Interaction Process Analysis. Cambridge, Mass.: Addison-Wesley Press, Inc., 1950.

2. Adler, L.M. and Enelow, A.J.: An instrument to measure skill in diagnostic interviewing: A teaching and evaluation tool. J. Med. Educ. 41:281, 1966.

3. Helfer, R.D., and Hess, J.: An experimental model for making objective measurements of interviewing skills. J. Clin. Psychol. 26:327, 1970.

4. Andrew, B.J.: A technique for the assessment of pharmacy students' skills in patient interviewing. Am. J. Pharm. Educ. 37:290, 1973.

5. The continuing search for betters ways to assess clinical competence. Natl. Bd. Examiner 21:No. 2, 1973.

6. Barrows, H., and Abrahamson, S.: The programmed patient: A technique for appraising student performance in clinical neurology. J. Med. Educ. 39:802, 1964.