INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106
Hixon, Sally J.

AN INVESTIGATION OF THE PSYCHOMETRIC PROPERTIES OF A CLINICAL SIMULATION EXAMINATION FOR RESPIRATORY CARE PRACTITIONERS

The Ohio State University

Ph.D. 1985

University Microfilms International

Copyright 1985 by Hixon, Sally J.

All Rights Reserved
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark (✓).

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received
16. Other Appendices have very poor and indistinct print. Best available copy.

Filmed as received.

University Microfilms International
AN INVESTIGATION OF THE PSYCHOMETRIC PROPERTIES
OF A CLINICAL SIMULATION EXAMINATION
FOR RESPIRATORY CARE PRACTITIONERS

DISSERTATION

Presented in Partial Fulfillment of the Requirements
for the Degree Doctor of Philosophy
in the Graduate School of the Ohio State University

By
Sally J. Hixon, B.S., Ed.M.

* * * * *

The Ohio State University
1985

Reading Committee:
William E. Loadman
Milton D. Hakel
Thomas E. Nygren

Approved By:
William E. Loadman
Advisor
Educational Theory and Practice
Note: This research project was not sanctioned by, nor were the contents of this manuscript approved by, The National Board for Respiratory Care, Inc.; this research was conducted for academic purposes only, and the contents of this report as well as the opinions expressed herein, are the sole responsibility of the author.
The purposes of this research were twofold. The first was to determine the dimensionality of the twenty problem score variables derived from a nationally standardized written clinical simulation examination (CSE), consisting of ten branching patient management problems. The second was to investigate the relationship between quantified measures of work experience and test scores.

Data was obtained for three examinee samples; two were representative of the candidate population eligible for the clinical simulation examination (CSE) and each completed the test for credentialing purposes (n=2821 and 2546), while the third sample (n=312) completed the test solely for research purposes and was representative of the U.S. practitioner labor force including subjects not eligible in terms of test admission requirements. As the CSE was designed according to the two factor cognitive process model demonstrated by factor analyses of
physician patient management problem tests, confirmatory factor analysis was used to assess the fit of the twenty score intercorrelation matrix to this model. The goodness-of-fit was determined to be unacceptable for one of the candidate samples as well as the research sample, both whom completed the same test form. Therefore, exploratory factor analyses were conducted for these two data sets and for the other sample of candidates who completed a different form of the test. Results of the principal axis factor analyses demonstrated remarkably similar one factor solutions for each sample, accounting for 80-86% of the variance. These results demonstrated the CSE to be unidimensional, and it was furthermore concluded that the one factor solution was stable and generalizable across samples of subjects and samples of problems comprising an examination.

Of the 312 research subjects provided a Work Experience Inventory (WEI), 112 were returned. A CSE factor score was derived to serve as the dependent variable, and results of an analysis of variance by WEI job level demonstrated a significant difference (p< .01) between subjects categorized as holding an advanced versus an entry level job. Additionally, three continuous WEI measures were determined to be significantly correlated with the CSE factor score (p< .01), indicating that experience and test performance are positively related.
To James and Audrey, my parents, who so graciously bestowed on me the Bonner and Hemphill heritage, taught me all that they could, provided me every opportunity in life, and always hoped only for my happiness in return.
ACKNOWLEDGEMENTS

I first wish to acknowledge the assistance of my friend and colleague Dr. Ann Schreck, who contributed to the success of this endeavor in many ways. Ann created the chapter one figure artwork, provided moral support and lodging, assisted me in wading through system commands so the university computer could read my data tapes, and provided much inspiration by completing her dissertation while I was "lost in the woods" with mine.

Next, I wish to thank Dr. William Loadman, my academic advisor and just at the right times, friend or mentor, but always a gentleman and a scholar. I respect Bill for his knowledge of testing and measurement, admire his competency in research methodology, and thank him for teaching me most of what I learned about these subjects at OSU. I also wish to thank Dr. Milton Hakel, who helped increase my awareness and understanding of the employer and legal perspectives regarding selection procedures in general, and job competency testing in particular. In addition, I wish to thank the remaining member of my committee Dr. Thomas Nygren, who provided me thoughtful questions on the statistics portion of my qualifying examinations, and consequently taught me much which helped to make this a solid quantitative study.

I also wish to thank the members of The National Board for Respiratory Care, who have helped me to grow personally and professionally over the years, and provided me many opportunities to do so. There are far too many to name, but they know who they are and how appreciative I remain. Without the interest of the National Board for Respiratory Care in general, this research would not have been possible; but without the efforts of Mr. Gary Smith and Mr. Steven Bryant specifically, it may not have been successful.

Additionally, I wish to thank Mr. Harold Levine who first introduced me to clinical simulations many years ago, as well as my current colleagues Maree Bellchambers and Dr. Brian Davis who provided me with intellectual stimulation and encouragement just when I needed it. Last but far from least, I thank Patrick Cavanaugh, my "significant other" in graduate school lingo, who fulfills my life with a special friendship and love.

iii
VITA

Professional Education and Experience

1973 A.A.S. in Respiratory Therapy, Upstate Medical Center, Syracuse, New York.


1975-78 Assistant Director, Respiratory Care Division, The Children's Hospital of Buffalo, New York.


1979-82 Director of Clinical Education, Erie Community College, Buffalo, New York.

1980-84 Clinical Simulation Examination Consultant, The National Board for Respiratory Care, Inc.

1982-84 Graduate Teaching Associate, Education, The Ohio State University, Columbus, Ohio.

1984 Declared to Ph.D. Candidacy, Doctoral Studies in Research and Evaluation, Department of Educational Theory and Practice, College of Education, The Ohio State University, Columbus, Ohio.

Publications


# TABLE OF CONTENTS

DEDICATION ................................... ii

ACKNOWLEDGEMENTS .......................... iii

VITA ............................................ iv

LIST OF TABLES ................................ viii

LIST OF FIGURES . ............................. ix

I. INTRODUCTION ............................. 1

- Background to the Problem ............... 4
- Statement of the Problem ................. 9
- Rationale and Purpose of the Study ...... 13
- Conceptual Framework ....................... 18
- Research Questions ......................... 27
- Importance of the Study .................. 28
- Limitations of the Study ................. 29
- Definition of Terms ....................... 31

II. REVIEW OF THE LITERATURE ............ 34

- Competency Assessment in the Health
  Professions ................................ 35
- The Criterion Problem in Evaluation ..... 39
- Assessing Competence with Simulations ... 42
- Assessing Clinical Problem Solving Ability . 46
- Credentialing in Respiratory Therapy . . 55
- Validity of Written Clinical Simulations . . 58
- Reliability and Generalizability .......... 58
- Criterion-related Validity Research ..... 64
- Research on Construct Validity .......... 67
- Theories of Human Problem Solving ...... 78
- Research in Psychology .................... 79
- The Medical Inquiry Project ............... 81
- Process Theories of Medical Problem Solving. 86

Chapter Summary ............................ 92
APPENDICES

A. Sample CSE Score Report ................................ 198
B. CSE Test Matrix - June 1984 Test Form ................ 200
C. CSE Sample Patient Management Problem ............... 202
D. Work Experience Inventory (WEI) Instrument ........... 209
E. Sample WEI Profile (Score Report) ...................... 223
F. Correlation Matrices for the Candidate and Study Data Sets . 226
LIST OF TABLES

1. Respiratory Therapy Competency Assessment System . 57

2. Description of Participating Hospitals . . . . . 101

3. Demographics of the Total Study Sample and the Subset Returning the WEI. 104

4. Candidate and Study Sample Descriptive Statistics . 129

5. MANOVA by Group for the Twenty Problem Score Variables . 131

6. Results of Each One-Way ANOVA by Group for the Twenty Problem Score Variables . 132

7. LISREL Chi-Square Goodness-of-Fit Values for the Candidate and Study Data Sets . 140

8. Rho Statistic Goodness-of-Fit Values for the Candidate and Study Data Sets . 141

9. Eigenvalues and Percent of Variance Explained for the Candidate and Study Data Sets . 145

10. Factor Loadings of the One Factor Solution for the Candidate and Study Data Sets . 146

11. Descriptive Statistics for the WEI Respondent Sample . 150

12. Results of the One-Way ANOVA for CSE Score by Entry and Advanced Job Level . 152

13. Intercorrelations for the Three WEI Score Variables and the CSE Factor Score Variable . 154

14. Results of the Multiple Regression Analysis of WEI Section Scores on the CSE Factor Score . 155

viii
LIST OF FIGURES

1. Application of the Bashook Clinical Domain Framework to the NBRC Clinical Simulation Examination . 22

2. Application of the Bashook Clinical Domain Framework to the Respiratory Therapy Work Experience Inventory . 24

3. SAS Scree Plot of Eigenvalues for the Candidate Sample Data Set . 142

4. SAS Scree Plot of Eigenvalues for the Study Sample Data Set . 143
I. INTRODUCTION

The efficient and effective evaluation of clinical competence in the health professions has been a long standing concern of educators as well as others (Woolley, 1977). Questions regarding the validity of such competence evaluation have prompted much research in the last decade (LaDuca, 1980; McGaghie, 1980). During this period, legal challenges pertaining to licensure, certification, and other assessment methods which fall under the rubric of selection procedures, have also focused attention to this issue (Bryant, 1981; Shimberg, 1982).

Central to this validity issue is the concept of competence and the difficulty in defining it for assessment purposes (McGaghie, 1984). Competence may be broadly defined as the ability to perform certain behaviors when presented with a particular set of conditions (Segall, Vanderschmidt, Burglass and Frostman, 1975). Clinical competence consists of unique combinations of skills, knowledges and attitudes necessary for safe and effective practice in a given position, and may vary from simple behavioral responses with few conditional
alternatives to complex integrations of unlimited scope (Scanlon, 1978). In psychological terms, competence is a multidimensional construct (McGaghie, 1980), and it is generally agreed that no single test method is capable of providing a reliable and valid assessment of all the components which constitute clinical competence (Small, 1975; Marshall, 1977; McGuire, 1963, 1969; Newble, 1979).

With any testing procedure, validity depends upon the interpretation or significance that is attributed to the test score (Cronbach, 1971). McGaghie (1980) states that "accurate personnel decisions are the goal of professional competence evaluation". He claims this is so because such decisions serve to protect the consumer from unqualified practitioners and enforce practice standards. Consequently, if an employer infers that attainment of a credential signifies competence, then the criterion for competence as assessed by the test becomes central to the validity issue. McGaghie (1984) has referred to this as the criterion problem in professional evaluation and Forsythe (1984) has noted that score interpretation is part of this problem.

In particular, traditional objective cognitive knowledge tests predominantly measuring recall (i.e., multiple-choice type), have been questioned in terms of providing a valid competency assessment, as the mere
possession of such knowledge does not render a practitioner clinically competent (Bashook, 1976; Lloyd, 1982; Martin, 1975; McGuire, 1963, 1969). It is recognized that such tests neglect other important aspects of competence and this has lead to the development of alternative methods to either supplement or replace traditional objective cognitive knowledge tests (McGuire, 1969).

Clinical judgment is an element of health practitioner competence considered to be central to effective performance (Williamson, 1965). This is commonly described, as well as operationalized for teaching and testing, as clinical problem solving ability in patient management (Andrew, 1972; DeTornyay, 1968; Dielman, Hull and Davis, 1980; LaDuca, 1979; Martin, 1975; McGuire, 1976; Vu, 1980). This component of competence is recognized as involving more than the mere mastery of medical knowledge (Donabedian, 1976), but is also noted to be difficult to assess (Marshall, 1977). Various forms of simulation are used in the health professions for teaching and testing clinical problem solving skills (Lloyd, 1982; McGuire, 1976; Scanlon, 1978). Kacmarek, Hixon, and Assmann (1985) provide the following summary of the rationale and purpose for the use of simulations in clinical education and evaluation.
To view competence as merely the technical mastery of subject matter is to reduce the complexity of professional competence to a very narrow focus. Beyond such mastery is the ability to understand concepts and complex relationships, synthesize information, render judgments, effectively communicate with patients and co-workers, make decisions and resolve problems. This is the level of complexity present in actual clinical practice. Because reality is so complex, so too it is difficult to provide for instruction and evaluation methods that adequately reflect this complexity, yet allow for the structure and standardization necessary for teaching and testing. One alternative to traditional instruction and evaluation methods that is useful for bridging this gap between theory and practice is simulation technique. (p. 2)

Background to the Problem

In 1978, written simulations for assessing clinical problem solving ability were developed for the field of respiratory therapy as one part of a federally funded project (American Association for Respiratory Therapy, 1978). The written clinical simulations devised were patterned after a type of instrument earlier proposed by Christine McGuire and her colleagues called a patient management problem. The National Board of Medical Examiners (NBME) had replaced the bedside practical portion of the physician credentialing process with a type of written simulation examination in 1961 (Hubbard, 1978), and later similar physician examining bodies in Canada and Australia adopted the technique. In 1979,
The National Board for Respiratory Care (formerly the National Board for Respiratory Therapy), replaced the unstructured oral portion of their advanced level credentialing examination with a written clinical simulation examination consisting of ten patient management problems.

A patient management problem (PMP) begins with a brief scenario which describes the clinical setting and initial presenting condition of the patient (McGuire, Solomon and Bashook, 1976). Various sections of the problem then provide a series of options concerning history, physical examination, laboratory studies, and treatment alternatives. Responses to the options are hidden by invisible (latent image) ink. An examinee records selections by development of the ink with a special pen, thus revealing information requested, results of an action, or direction to a subsequent section of the problem (Assmann, Hixon and Kacmarek, 1979).

A PMP may be simply described as a series of interrelated items in which feedback about the consequences of each decision is provided before one proceeds to the next section of the problem (Sedlecek & Nattress, 1972). Design characteristics, however, determine the order in which sections of the problem are completed. The branching type of PMP allows for multiple approaches to problem completion as the order in which sections are
completed varies based on selections made at divergent points in the problem; thus examinees do not complete all sections (McGuire, 1976). On the other hand, the linear type of PMP presents the sections for examinees to complete in a prescribed order (Marshall, 1983).

Prompted by the use of PMPs in credentialing, most medical schools and many health practitioner training programs utilize written simulations for teaching and/or testing clinical problem solving skills. However, many of these "teacher-made" instruments lack the rigorous developmental effort devoted to credentialing examinations and differ widely in format, stimulus and response conditions, and scoring procedures (Vu, 1979). A review of the literature by Galofre in 1974 revealed eighteen variations of the PMP (Newble, 1979).

Despite the attempt to delimit clinical competence to problem solving ability, and to assess such with written clinical simulation tests, there is a great deal of controversy regarding the reliability and validity of these measures (LaDuca, 1980; Swanson, 1984). Due to the great diversity of instruments used for research in this area, as well as different scoring and analytic procedures, it is not surprising that results have varied (Swanson, 1984).

Several studies investigating the validity of PMPs in terms of comparing test performance to behavior in
actual practice (as recorded in patient records) or to other forms of free-response simulation (such as oral or short answer), have demonstrated that there is a significant "cueing" effect in PMP performance (Blumberg, 1981; Goran, Williamson and Gonnella, 1973; McCarthy, 1966; Newble, Hoare and Baxter, 1982; Norman and Feightner, 1981; Page and Fielding, 1980). The consistent finding is that written simulation formats which "cue" the examinee by providing a list of options to choose from, result in more data collection thus affecting problem solving behavior. It seems clear that abilities assessed by a simulation test differ based on the type of instrument used (Scanlon, 1978).

Even when studies have used PMPs in the structured response format proposed by McGuire, Solomon and Bashook (1976), methodological problems still arise because items in a PMP differ in many important ways as compared to traditional objective cognitive knowledge tests such as: (1) items are not independent but rather are interrelated in a PMP so performance on one affects performance on others, (2) options are not simply determined as right or wrong (i.e., 0/1 weighting), but are differentially weighted according to their relative contribution to resolving the problem and include negative weights, and (3) in a branching PMP, all examinees do not complete the same set of items (McGuire, 1969).
Consequently, it has been argued that traditional concepts of reliability and validity are not applicable to PMPs (Dincher and Stidger, 1976; McGuire and Babbott, 1967; Sedlacek and Nattress, 1972; Small, 1975).

Such measurement problems have led to an emphasis on construct validation research to clarify the nature of problem solving and its measurement. Indeed, construct validation is particularly important when assessment criteria consist of hypothesized abilities which underlie test performance (Forsythe, 1984), and involve a patterned set of interrelationships (LaDuca, 1980). Most of the research in this area has been restricted to medical problem solving using samples of physicians and medical students, and written simulation tests which emphasize diagnostic skills (Vu, 1980). The fundamental controversy that has arisen from this research regards the extent to which medical content, as opposed to a general problem solving process, underlies clinical problem solving ability (Harasym, Alexander, Baumber, Bryant, Fundytus, MacPhail, Preshaw, Sosnowski, Watanabe and Wyse, 1979). That is, some research results indicate that clinical problem solving is primarily dependent upon medical content specific within a case, while others indicate that it is a skill or series of skills which can be applied to all clinical problems.
Statement of the Problem

The case specificity theory of medical problem solving resulted from the Medical Inquiry Project directed by Arthur Elstein and conducted at Michigan State University (Elstein, Shulman and Sprafka, 1978). One portion of this large research project analyzed the performance of groups of physicians and medical students on three PMPs using single summary scores for each problem. A great deal of intergroup and intraindividual variation in performance was discovered across problems which was interpreted as signifying that problem solving strategies were specific to the nature of the problem posed and the relevant experience of the subject. These results were confirmed by verbal stimulated recall sessions following simulated videotaped encounters with actors trained to play patients. In fact, much earlier McGuire and Babbott (1967) had noted this variation phenomenon and had strongly suggested that it was necessary to differ the context of patient situations across problems so that one approach to problem solving is not consistently rewarded. In a later article however, McGuire (1976) interpreted low correlations between PMP scores as an indication that the problem solving skills measured were case specific.

Discovery of this phenomenon is not too surprising if one considers research in cognitive psychology which
characterizes the process of human problem solving as adaptive in that problem solving behavior is determined by the demands of the task environment (Simon and Newell, 1971). Human problem solving is perceived as a person-environment interaction (Hunt, 1975), and therefore the structure of the tasks posed to the problem solver may be as important in the study of problem solving behavior as the process the subject uses (Simon and Newell, 1971).

Many authors have refuted case specificity findings as a methodological artifact (Berner and Engel, 1979; Harasym, Baumber, Bryant, Fundytus, Preshaw, Watanabe and Wyse, 1980; LaDuca, 1979, 1980; Marshall, Fleming Heffernan and Kasch, 1982; McGaghie, 1979; Vu, 1979, 1980). Common criticisms are that the Medical Inquiry Project studies consisted of in-depth investigations of few subjects and utilized a small number of problems that were selected arbitrarily in reference to content (LaDuca, 1980). Therefore, the generalizability of these research results is suspect primarily because the simulations failed to represent the environment of internal medicine clinical situations and therefore lacked content validity (LaDuca, 1980). Additionally, the relationship between the subject and the clinical problem could not be addressed because the focus of the research was not on that relationship but rather on the
description of an assumed underlying cognitive trait (LaDuca, 1980).

Studies which have used large subject samples and aggregates of ten or more PMPs have produced quite different results, but they have also used different scoring procedures and analytic techniques. Donnelly, Gallagher, Hess and Hogan (1974) conducted a factor analysis of performance measures from 162 medical students on ten PMPs. A separate score was derived for seven content areas represented in each problem which corresponded to the various stages in the diagnostic work-up and treatment of a patient. Many factor analyses were performed - one for each of the ten individual problems and one for the aggregate using the average score across all problems for the seven content areas. The aggregate analysis demonstrated a clear two factor structure, as did eight of the individual problem analyses. Factor I was defined as a linear combination of the history, physical exam, laboratory and radiograph section scores and called information gathering; factor II was termed decision making and defined as a linear combination of the diagnosis, pathway and management scores. Additionally, the canonical correlation between the average scores for the variables constituting each factor demonstrated that they were not significantly correlated. It was concluded that two scores are
required to adequately describe examinee performance on a PMP and since problems are not unidimensional, it was advisable to determine PMP reliability in regard to a given dimension across problems rather than by problem.

A similarly conducted study by Juul, Noe and Nerenberg (1979) involved three large groups of medical students and factor analyses of six average content area scores derived from tests composed of more than twenty PMPs. The results confirmed that of Donnelly et al. (1974) demonstrating a parallel two factor structure, and the factors were shown to be stable for different groups taking the same examination and over time for one group who took two different forms. This provided further evidence in support of the theory that two components constitute medical problem solving as measured by PMPs.

Regarding the nature of medical problem solving, McGuire (1976) suggested that the apparent content dependence of these skills may be an artifact in the variation of format of the simulations used, a consequence of a fragmented disease-oriented medical instruction system, or a very real fact of life. Certainly, effective problem solving is not possible without some degree of knowledge relative to the content of the problem, and many authors have noted the importance of experience in the substantive domain for which problems are to be
solved (DeGroot, 1965; Elstein, Shulman and Sprafka, 1978; McGuire, 1976; Simon and Newell, 1971). However, the debate relative to process versus content in medical problem solving remains unresolved and despite the use of PMPs to assess the clinical problem solving skills of allied health practitioners such as respiratory therapists, no research has been conducted outside the domain of physician practice relative to identifying cognitive processes which underlie performance on PMPs.

Rationale and Purpose of the Study

The purpose of this study is to investigate the psychometric properties of a written clinical simulation examination designed to assess the problem solving ability of respiratory care practitioners. This examination consists of ten branching patient management problems and is currently used for credentialing respiratory therapists in the United States by the National Board for Respiratory Care (NBRC).

The NBRC Clinical Simulation Examination (CSE) consists of ten branching patient management problems, and is designed specifically for assessing the clinical competence of the advanced level respiratory therapist (National Board for Respiratory Care, 1984). The CSE is designed to assess this level of practice based on a
Candidates for the CSE must have successfully completed the entry level multiple-choice examination which is taken upon graduation from an AMA accredited respiratory therapy program, and have completed one year of documented clinical experience following graduation to be eligible. The following information describing the CSE is excerpted from the NBRC study guide for the advanced level credentialing process (NBRC, 1984).

The purpose of the Clinical Simulation Examination is to assess the clinical problem solving ability of the candidate, including identifying and analyzing the problem, and taking appropriate steps toward its solution. The problem solving model incorporated in the Clinical Simulation Examination is as follows:

1. Identifying the Problem - Information gathering sections are designed for this purpose ... physical and laboratory data obtained provides the basis for subsequent decisions in patient care and management ... Candidates are expected to be selective, but thorough, in gathering pertinent data ... for rendering sound clinical judgments. Candidates may not be successful in these sections if they select an insufficient amount of data to safely and effectively substantiate a conclusion. They may also be penalized for selecting unnecessary options.

2. Analyzing the Problem - This consists of interpreting and evaluating data ... This step in problem solving is not directly assessed ... (but) is a process you must use to arrive at a decision or judgment regarding the care of the patient.
3. Resolving the Problem - Decision making sections are designed for this purpose ... candidates are required to render clinical decisions relative to the care of the patient with regard to timing and appropriateness of therapy. (p. 34)

The salient assumption of this model is that clinical problem solving is predominately a process activity, presumably considered common to all cases. Moreover, although individual problem and content area scores are generated for candidate feedback on performance, only two composite scores are used to determine pass/fail status. One is an information gathering (IG) score, which is merely the sum of scores from all sections designated as such; and the second is a decision making (DM) score which is a similarly summed value from all these type of sections. The candidate must achieve a pre-established passing level in both categories to pass the examination. A sample CSE score report is presented in Appendix A, and the design and scoring of the CSE is discussed further in Chapter Three.

The content validity of the CSE can be asserted as job analysis data forms the basis for exam content, and standardized procedures are utilized in developing problems and constructing test forms (Hixon, 1982). A confidential criterion-related validity study has been conducted for the CSE, demonstrating statistically significant correlations between the criterion measure
(supervisor ratings of job performance) and the IG and DM scores respectively (Bryant, 1985). However, these two scores have never been empirically established as representative of dimensions of test performance, and the relationship of experience to test performance has not been studied despite the inclusion of this requirement for admission to the examination. Considering the conflicting research results regarding the nature of medical problem solving and measurement with PMPs, an investigation of the dimensionality of measures derived from the CSE and the relationship of work experience to test performance seems warranted.

Additionally, the NBRC written simulation exam differs from those used for physician competency assessment in several ways. Central to these differences is that the domain of practice for respiratory therapists is much more discrete than that of physicians; in the former, boundaries are more easily defined and job tasks have been delineated by systematic national studies of roles and functions (American Association for Respiratory Therapy, 1977; Nettles, 1982). Also, physician PMPs commonly emphasize the process of diagnosing patients with a wide range of health problems, while respiratory therapy PMPs are most concerned with technical (i.e., therapeutic) skills necessary for clinically managing patients with pulmonary disorders.
Moreover, sections of the NBRC problems are designated as IG or DM on the basis of item type rather than on stages in the clinical work-up and care of patients as they are in physician PMPs, and derived summary scores differ accordingly. That is, IG sections in NBRC problems consist of lists of approximately 15-30 information options from which the candidate selects as many as are considered indicated, whereas DM sections typically offer 5-8 decision options from which the candidate selects only one (National Board for Respiratory Care, 1984). Total IG and DM scores are determined by summing weights assigned to the options selected in each of these item types, and in part the rationale for deriving two scores was that the greater number of options offered in IG sections would act to "contaminate" what is perceived as the more important component of rendering clinical decisions regarding patient management (Kacmarek, Hixon and Assmann, 1985). Therefore, despite the superficial similarity of physician and respiratory therapy PMPs, differences in the focus, content and format of the problems, as well as derived scores exist. The effect of these differences relative to the nature of clinical problem solving as measured by PMPs is unclear.
Conceptual Framework

The complex and elusive nature of competence renders it difficult to define and delineate a criterion for professional competency assessment (McGaghie, 1984). Competence is a multidimensional construct which can not be observed directly (McGaghie, 1980); rather indicators of competence must be delineated, systematically observed, and inferences advanced based on these observations (Cronbach, 1971). Consequently, many credentialing bodies utilize multiple indicators in order to assess competence and render decisions regarding the credentialing of health practitioners (Marshall, 1977). Therefore, the issue becomes one of explicating the meaning of measures derived from observation of competence indicators (Messick, 1975).

On the one hand, we have the criterion of medical knowledge which is regarded as a necessary but insufficient indicator of competence (Lloyd, 1982). Objective cognitive knowledge tests are commonly used to assess this aspect of competence. These tests are generally quite reliable but lack adequate validity to represent the global criterion of competence (McGuire, 1963).

On the other hand, we have the criterion of clinical problem solving ability which is perceived as a higher order cognitive process that requires an integration of medical knowledge and clinical experience in
managing patients (Donabedian, 1976). Written clinical simulation instruments provide a standardized and objective method for assessing this aspect of competence (Kacmarek, Hixon and Assmann, 1985), and performance on such tests is logically conceived as a better indicator of the global criterion of competence because of the similarity of tasks on the test and tasks in clinical practice (McGuire, 1976). However, because of the differences in item characteristics of such tests (i.e., interdependence and differential weighting), the use of traditional methods for estimating reliability and validity are not unequivocally applicable (Small, 1975), and a great deal of controversy exists regarding the psychometric properties of these instruments. In other words, the meaning of measures derived from written simulation instruments is unclear and requires explication.

The many differences in the design, format, focus and scoring of written simulation instruments used for research in this area make it difficult to compare results and draw meaningful conclusions regarding the measurement of clinical problem solving ability (Swanson, 1984). Therefore, a systematic approach for considering the psychometric properties of a written simulation instrument is required in order to derive meaningful conclusions regarding the nature of that which is measured (Vu, 1979).
Bashook (1976) suggests that the conflicting data from studies of physician PMP performance indicates the necessity for defining the domain in which the clinician functions as the starting point for operationalizing clinical problem solving. He argues that attempts to measure clinical problem solving ability have focused on modeling an unknown but presumed underlying process and have neglected sufficient consideration for the various types and scope of clinical problems a practitioner is required to solve. Bashook (1976) presents a conceptual framework for measuring clinical problem solving ability which consists of the following three dimensions: (1) problem-solving process, (2) clinical discipline, and (3) context of care. The intersection of these three dimensions defines the clinical practice domain providing a framework for determining the number and type of patient problems which must be included in a clinical simulation examination in order for the content of the test to be considered representative of the clinician's domain of practice.

The NBRC clinical simulation examination can be perceived in terms of this three dimensional framework and shown to be representative of the respiratory care practice domain. First, the problem solving model for the assessment is delineated as a three step process with two steps operationalized for measurement, thus
defining dimension one. Second, the content of the exam is based on a national job analysis which delineates the job tasks and duties of the advanced level practitioner. For purposes of the CSE test matrix, clusters of related job tasks have been categorized into thirteen content areas, and the sections of each PMP are identified as belonging to one of these content areas relative to item type (IG or DM). A copy of the test matrix for the CSE form to be used in the study may be found in Appendix B. The ten problems that comprise any one form of the exam are selected such that the five content areas which are amenable to IG assessment, and each DM content category are proportionally sampled. Thus the second dimension, clinical discipline, is defined and represented.

Additionally, problems are categorized in terms of type of patient (i.e., neonatal, pediatric or adult), type of disease (i.e., cardiovascular, neuromuscular, acute and chronic pulmonary), and setting (i.e., general floor, emergency room, intensive care unit). The aggregate of problems selected to comprise any one form of the exam must meet preset scope of practice criteria such that the range of contexts of care are represented, thus defining dimension three. Figure 1 on the next page illustrates an application of the Bashook clinical domain framework to the NBRC Clinical Simulation Examination.
Figure 1. Application of the Bashook Clinical Domain Framework to the NBRC Clinical Simulation Examination
A second instrument, the Work Experience Inventory (WEI) for Respiratory Therapists, has been developed which provides alternative measures of clinical abilities in this domain. Modeled after those developed for quantifying the work experience of applicants to graduate schools of management, the WEI was developed and pilot tested by Educational Testing Service for the NBRC (Nettles, 1984). This self-report instrument was designed to provide a standardized method for quantifying the relevant work experience of practicing respiratory therapists. The instrument consists of many sections; each provides measures of work experience which are perceived to represent clinical abilities.

The first section of the instrument consists of nine weighted items which are collectively referred to as Job Factors. Total points achieved in this section indicate the level of responsibility in a job and allow for classification of the job as entry, advanced or administrative. This may be viewed as a summary measure which provides for classification, much like the total IG and DM scores for the CSE summarize overall performance and allow for classification (pass/fail status).

The remaining three sections of the WEI may be perceived as representing the three dimensions of Bashook's clinical domain framework. This is illustrated in Figure 2 on the following page.
WORK EXPERIENCE INVENTORY

I. Problem Solving Process
Knowledges, Skills and Abilities

II. Clinical Discipline

III. Context of Care
Work Situations
Position Responsibilities

Figure 2. Application of the Bashook Clinical Domain Framework to the Respiratory Therapy Work Experience Inventory
The WEI Knowledges, Skills and Abilities section consists of sixty-two items which provide a measure of the significance of these elements to job performance. As medical knowledge is considered integral to problem solving ability, this section of the WEI is considered related to the problem solving process dimension.

The WEI Position Responsibilities section consists of eighty-two items which indicate the scope of job tasks and duties performed. These represent a subset of the job analysis items which form the basis for the thirteen content areas identified on the CSE matrix, and therefore are considered to define the clinical discipline dimension.

The Work Situations section consists of twenty-two items which indicate the breadth of clinical exposure in the field. These items present work situations with various types of patients, settings of care, and roles in providing care. Thus, this section is considered to represent the context of care dimension.

Viewing the CSE and WEI in terms of Bashook's three dimensional clinical domain model provides a framework for the study of the relationship between the subject (person) and his/her relevant domain of practice (environment), for each instrument individually as well as for the comparison of measures from both. As the problems comprising any one form of the CSE are not
an arbitrary aggregate but rather can be demonstrated as sufficiently representative of the relevant aspects of respiratory care clinical practice, an analysis of the dimensionality of measures derived from this examination should provide meaningful information relative to the case specificity versus process controversy. Also, as the WEI is a comparable instrument in terms of representing the relevant aspects of this clinical domain, the WEI scores provide measures which can be compared to CSE performance measures in order to determine the relationship between clinical experience and problem solving ability. Specific research questions for the study follow on the next page.
Research Questions

Several questions regarding the assessment of clinical problem solving ability, and the psychometric properties of the CSE, guide this research effort. The first set of questions addresses the case specificity versus process controversy and involves an analysis of the dimensionality of measures derived from one form of the CSE. The remaining questions are directed at investigating relationships between CSE test performance and measures of relevant work experience.

1. What is the goodness-of-fit of the IG and DM problem scores to the hypothesized two factor model?

(a) If the fit to this model is poor, what is the underlying dimensionality (i.e., factor structure) of the CSE problem scores?

(b) Whatever the results of the above, can similar results be demonstrated for a different sample of examinees who have completed the same CSE test form (i.e., can the model be generalized to other samples)?

2. Do statistically significant differences in CSE test performance exist between groups of practitioners classified as entry, advanced and administrative on the basis of the WEI Job Factors score?

3. What is the direction, magnitude and statistical significance of the relationship between CSE test performance and the:

(a) WEI Work Situations score?
(b) WEI Position Responsibilities score?
(c) WEI Knowledges, Skills and Abilities score?
Importance of the Study

The construct component of validity is central to the interpretation of scores produced by a set of measurement operations, and the process of construct validation is directed at determining the meaning of what is measured (American Psychological Association, 1984). The construct should be defined by a conceptual framework which indicates how observed measures should be interrelated, and relatively correlated or uncorrelated with other measures (American Psychological Association, 1984). The psychometric properties of the CSE have not been systematically studied within such a framework. The assumption that IG and DM processes reflect clinical problem solving ability, and that measures derived from these dimensions are indicative of professional competence in terms of work experience, have not been empirically tested. These assumptions are central to the design of this test as well as the interpretation of scores.

Selection and promotion decisions are often made, at least in part, by the results of credentialing examinations (Bryant, 1981). Although the attainment of an NBRC credential is not required by law for the practice of respiratory therapy, such voluntary credentialing can be equally as restrictive as licensure when employers choose to use the results of this testing for employment
decisions (Shimberg, 1982). Consequently, a practitioner's career opportunities may be determined by the interpretation of these test scores, which in turn may ultimately affect one's livelihood.

On the other hand, efficient employment decisions are important for consumer protection (i.e., patient safety) and with health care costs rising, also vital to the economy. Therefore, substantial justification exists for improving our understanding of professional competence in general, and clinical problem solving ability in particular. In order to effectively teach, evaluate and provide corrective feedback to future health care practitioners, an understanding of the ways in which information is utilized to identify and resolve clinical problems is necessary (Helfer and Slater, 1971). Beyond these pragmatic concerns lies the importance of contributing to the theoretical understanding of this poorly defined and perplexing set of cognitive processes involved in problem solving ability (Elstein, Shulmam and Sprafka, 1978).

Limitations of the Study

One of the limitations of the study involves the representativeness of the sample. Random sampling of respiratory care practitioners is not possible because
no comprehensive listing of practicing respiratory therapists in the nation is available, and supervised group administration of the CSE is required with standardized testing conditions.

A second category of limitations relates to the quality of data obtained in the study. Examination performance may be affected by a subject's understanding of test directions and familiarity or previous experience with this type of test. Also, bias may be introduced by the use of the WEI which is a self-report survey. Accurate responses on the survey will depend on the ability of the subject to understand the directions and items contained in the survey, to accurately recall their job responsibilities, and to properly select responses to reflect them. Also, all subjects may not respond to the survey which may introduce a response bias.

The interpretation of results presents a third limitation of the study. Interpretation of factor structures requires a substantive knowledge of the domain of variables and is highly subjective. Interpretation will be guided by theories of medical problem solving with particular attention to the process versus case specificity controversy. However, as problem solving performance may differ according to health specialty and the format of PMP used, conclusions may only apply to measures of problem solving ability for
the clinical domain of respiratory therapy and as assessed with the branching type of structured response patient management problem.

**Definition of Terms**

**certification** - the process by which a nongovernmental agency or association grants recognition to an individual who has met certain predetermined qualifications specified by that agency or association (U.S. Department of Health, Education and Welfare, 1977).

**clinical simulation examination (CSE)** - a written branching patient management problem test, employing the latent image response technique, designed to assess clinical problem solving abilities.

**competence** - ability to perform certain behaviors when presented with a particular set of conditions (Segall, Vanderschmidt, Burglass and Frostman, 1975).

**credentialing** - the formal recognition of professional or technical competence; a generic term referring to the processes of certification and licensure (U.S. Department of Health, Education and Welfare, 1977).

**decision making (DM)** - a component of clinical problem solving ability which involves rendering judgments and decisions relative to patient care and management.

**dimension** - magnitude measured in a particular direction; variable or set of homogeneous variables which describe a construct.

**information gathering (IG)** - a component of problem solving ability which involves the gathering of pertinent clinical data.

**latent image** - a technique by which responses to options on PMPs are rendered invisible; a candidate selects an option by using a special developer pen to expose the response.
licensure - the process by which an agency of government grants permission to an individual to engage in a given occupation upon finding that the applicant has attained the minimal degree of competency necessary to insure that the public health, safety, and welfare will be reasonably well protected (U.S. Department of Health, Education and Welfare, 1977).

patient management problem (PMP) - a pen and paper exercise designed to assess the clinical problem solving skills of health care practitioners; they may be of two design types:

linear - that type of PMP in which the sections are presented and completed in a prescribed order.

branching - that type of PMP in which the order sections are completed varies based on selections at divergent points in the problem.

problem solving - the cognitive process of identifying, analyzing, and resolving problems.

respiratory care practitioner / respiratory therapist - an individual responsible for providing respiratory therapy to patients in a clinical setting, such as a hospital; two levels of practitioners have been identified by national job analysis studies (Nettles, 1981; Nettles, 1982):

entry - that level of clinical practice or job responsibility characteristic of a practitioner just entering the field.

advanced - that level of clinical practice or job responsibility characteristic of accomplished practitioners in a field.

simulation - a representation of some aspect of the universe, designed to help us to understand, manipulate, enjoy, and/or predict the behavior of that universe (McGuire, Solomon and Bashook, 1976).

validity - refers to the appropriateness, meaningfulness and usefulness of specific inferences that are made on the basis of test scores; there are three kinds of validity evidence (American Psychological Association, 1984):
**content** - the extent to which the content of a test is representative of some universe of content.

**criterion-related** - the extent to which inferences based on test results are systematically related to one or more success criteria.

**construct** - extent to which the pattern of relationships between specific inferences from a test and other variables clearly indicate the meaning of the test score.

**work experience inventory (WEI)** - a self-report survey instrument which allows a practitioner to describe relevant aspects of work experience and provides for quantification of same (Nettles, 1984).

**written simulation** - a pen and paper exercise designed to simulate the decision making process; the individual is posed with a realistic problematic situation that requires a sequence of inquiries, decisions, and actions (McGuire, Solomon and Bashook, 1976).
II. REVIEW OF THE LITERATURE

Literature related to this study involves the fields of psychology, education and medicine. Relevant contributions from each of these literature bases will be included in the review which is organized into three major sections. The first section presents an overview of competency assessment in the health professions, including a discussion of the shortcomings of traditional test methods and the perceived usefulness of simulation technique for assessing aspects of competence which are difficult to assess by other means. This section concludes with a description of the nationally standardized competency assessment system currently in effect for the field of respiratory therapy in the United States. The second major section provides a review of the research on the reliability and validity of written clinical simulation instruments, thereby centering on measurement and instrumentation issues. The last major section of this chapter addresses the controversial issue of the nature of human problem solving and includes a description of opposing theories. The chapter concludes with a brief summary which focuses on the design of this study.
Competency Assessment in the Health Professions

The standardized mechanism for assessment of competence in the health professions is broadly referred to as credentialing, a term which encompasses the processes of voluntary certification and government regulated licensure (U.S. Department of Health, Education and Welfare, 1977). In order to attain a credential in most health occupations, an individual must first successfully complete a specified type and amount of educational training, then sit for and pass a standardized test as offered by a central agency (McGaghie, 1980).

The model for assessing competence in the health professions is rooted in medical education history. Medicine established a uniform educational system which provided the framework for a standardized mechanism of professional evaluation; hence the formation of the National Board of Medical Examiners in 1915 (McGaghie, 1980). Other health professions tend to follow this medical education and evaluation model in order to become recognized by, and established in, the professional health care community (Wilensky, 1964). An emerging health profession generally does this by developing relatively fixed systems of preprofessional training and establishing a national board which implements standardized methods for certifying or licensing practitioners.
The National Board for Medical Examiners (NBME) utilizes a three part testing process in which each test is designed to coincide with a particular stage in the physician training program in order to sequentially assess the competence of medical students prior to entering practice. Before the early 1950's, the Part I and II examinations were essay tests which each took three days to complete; as the emerging science of educational measurement demonstrated the merits of objective multiple-choice methods, the Part I and II examinations were converted to two day multiple-choice tests (Hubbard, 1978). However, Part III remained an unstructured bedside oral examination using actual hospital patients randomly assigned to examinees for approximately ten more years, despite dissatisfaction with its lack of objectivity (Hubbard, 1978).

Hubbard (1978) states that three functions have evolved on which the profession and public rely for "assurance that the (medical) education system is producing desired results." He enumerates these as: (1) evaluation of the student who is a product of the system, (2) evaluation of the educational process, and (3) evaluation of the institution responsible for the education. Presumably, only the first of these is directly assessed by the examination process, with perhaps some indirect evidence regarding the other two.
Nevertheless, by its design, the assessment system model established by the NBME clearly integrated competency assessment with the educational process. This implicitly gave rise to what may be called the "traditional" approach to competency assessment in the health professions.

The traditional approach to the development of such examinations, was to assemble a committee of "experts" in the occupation, who by consensus determined the content of the test (McGaghie, 1984; McGuire, 1969). Typically such tests followed a curricular model reflecting the scope of the required educational prerequisites, and were the objective cognitive knowledge type test (McGuire, 1963). There has been a growing dissatisfaction with this type of competency assessment method for many reasons (Swanson, 1984), but primarily because such tests tend only to provide an indication of what has been termed "academic" competence to the exclusion of providing any evidence relative to an examinee's ability to practice in the clinical setting (Kacmarek, Hixon and Assmann, 1985).

More than twenty years ago, medical education specialists began to challenge the traditional method of assessing competence, arguing instead for a "process approach" to the development of examinations (McGuire, 1963; Rimoldi, 1955). This approach called for emphasis
on the intellectual activity required to answer test questions, rather than concentration on covering curriculum in the "one right answer" type of test (DeTornyay, 1968; Rimoldi, 1960). The inadequacy of the traditional approach was exemplified when a systematic study of the 1961 NBME written examinations revealed that 93% of the items required only basic recall or simple comprehension abilities to answer correctly, and the form in which the item was posed had no relation to the intellectual activity required to answer it (McGuire, 1963). While theoretical knowledge certainly has a role in health practitioner performance (Small, 1975), it is generally agreed that the mere mastery of factual knowledge does not render a practitioner competent (Assmann, Hixon and Kacmarek, 1979; Bashook, 1976; Kacmarek, Hixon and Assmann, 1985; Martin, 1975).

Additionally, the alternate test methods used at the time, which were to provide a more relevant indication of an examinee's clinical abilities such as the NBME Part III bedside oral, were suspect because they were so advertantly unreliable. As Hubbard, Levit, Schumacher and Schnabel (1965) described the problem, the bedside evaluations were influenced by three variables - the examiner, the examinee and the patient; a reliable measurement of the variable of interest - the examinee, was not possible as the other two variables
could not be controlled. Therefore, the assessment of competence remained focused on performance on objective cognitive knowledge tests which were generally quite reliable, but insufficient as indicators of clinical competence.

**The Criterion Problem in Evaluation**

During the past two decades, controversy regarding the validity of credentialing examinations has continued within the professional evaluation field, and spread to the public arena as well. Public accountability has become the watchword (Hubbard, 1978), and validity the issue because the validity of a test is based on the interpretation or significance that is attributed to the test score (Cronbach, 1971).

The purpose of credentialing is to provide quality assurance, and the utility of the test score is determined by the accuracy of personnel decisions on which it is based (McGaghie, 1980). The measurement and recognition of individual competence is commonly justified on the grounds that it serves to protect the public from unqualified practitioners and assures professional autonomy (McGaghie, 1980). However, traditional test methods tend to measure only very limited aspects of competence to the extent that performance on typical tests bears little relation to effective practice.
(McGuire, 1969). Consequently, there has been considerable criticism that professional quality is not assured by such methods (Lloyd, 1982), and consumer demands for professional accountability have led to an emphasis on aspects of competence which are relevant and significant in actual clinical practice (Small, 1975). These concerns, as well as those dealing with equal opportunity in employment, were reflected in the adoption of federal guidelines in 1978 which advise that selection procedures used for personnel decisions be job-related and validated by job relevant criteria (U.S. Equal Employment Opportunity Commission et al., 1978).

Such issues of validity in credentialing have been collectively referred to as the criterion problem in professional evaluation, and described as "the challenge of identifying or describing the domains of skill or knowledge, professional qualities, sentiment and disposition, human affairs or any other facets that can form the basis of professional competence evaluation" (McGaghie, 1984, p. 3). The root of this problem is the difficulty with clearly defining the attributes underlying competence in a given profession, and providing for measurement techniques which link test performance with actual clinical performance. As McGaghie (1980) has described it, criteria refer to the "what" of evaluation, measurement to the "how", and greater conceptual
clarity of criteria is needed to sharpen test methods. Consequently, score interpretation is part of the criterion problem because inaccurate personnel decisions may result from test scores which are unrelated to actual practice (Forsythe, 1984; McGaghie, 1980).

Similar dissatisfaction with, and criticism of, traditional cognitive knowledge tests utilized in employee selection may be found in the personnel psychology literature of the same period. In a classic article by Wernimont and Campbell (1968), they proposed a "consistency model" for the development of job ability tests. This requires that a comprehensive study of the job be undertaken to define the criterion for testing, so that tests may be devised which focus on meaningful samples of behavior which are elicited under stimulus conditions like those on the job. In this way, measures can be obtained that are as similar to the criterion as possible. Such a test is designed to be a miniature replica of the job and is called a work sample test (Asher & Sciarrino, 1974). A work sample test incorporates dimensions of actual job behavior and emphasizes the contextual or situational factors deemed to influence job behavior. Asher and Sciarrino (1974) added to this consistency notion by proposing a point-to-point theory which states that complex work sample tests should be more valid and powerful predictors of job
performance because the predictor (test) shares many points in common with the criterion space (job performance).

Concerns regarding the criterion problem in professional evaluation have led to alternative conceptions of competence and development of innovative methods to assess aspects of competence which are not amenable to traditional types of objective cognitive knowledge tests (McGuire, 1969). Barro (1973) has stated that the "ultimate criterion" of clinical competence is actual practice and if the goals of professional evaluation are defined accordingly, it is possible to develop assessment methods that sample directly the behavior to be evaluated (McGuire, 1969). Simulation technique is one such method which provides for the assessment of clinical competence in a realistic manner.

**Assessing Competence with Simulations**

Clinical competence is a complex phenomenon and not amenable to simple solutions in terms of assessment (Newble, 1979). In psychological terms, competence is a multidimensional construct (McGaghie, 1980) and therefore can not be measured directly. The focus of competency assessment must consequently be on the identification and measurement of sound indicators of competence.

The complex interactions characteristic of clinical
performance are difficult to delineate, and therefore difficult to measure (Scanlon, 1978). In an attempt to make competency examinations more directly related to clinical practice, some examining bodies have undertaken systematic studies of health practitioner performance such as job analysis survey research, in order to provide a definitional basis for assessment (Newble, 1979). However, LaDuca (1980) has suggested that research methods which merely provide inventories of discrete bits of knowledge and skill may result in items which are more relevant to practice, but traditional test methods continue to measure behavior responses which are isolated from "real world" circumstances. He argues persuasively that the importance of performance is context-bound and competency assessment must provide a situational reference that accounts for the range and complexity of conditions surrounding job performance.

Simulation technique has gained widespread use in the health professions because it provides a method of representing life-like situations for evaluation. Reduced to its essence, simulation provides a method for imitating a clinical encounter and posing a problematic situation which requires a series of inquiries and decisions on the part of the examinee in an effort to resolve the problem (McGuire, 1976).

A wide variety of simulation formats have been
used, but each is basically one of four types:

(1) interactive simulator models such as the Resusci-Annie CPR mannequin;

(2) oral simulations such as the use of trained actors as patients, or the structured oral format wherein an examiner relays a patient situation and the subject's verbal responses are assessed;

(3) written simulations such as the patient management problem utilizing a latent image response technique; and

(4) computerized which are similar to the written type but are adapted for computer administration (Assmann, Hixon and Kacmarek, 1979).

All differ in terms of their application to various clinical skills (Scanlon, 1978). For instance, interactive laboratory models are best suited for assessing psychomotor skills, oral for communication skills, and written for cognitive skills involved in patient care. The types also vary in terms of their fidelity in representing reality, expense to administer, advantages and disadvantages relative to teaching versus testing, scoring methods utilized, and reliability as well as validity in evaluation (Assmann, Hixon, and Kacmarek, 1979; Swanson, 1984).

The written type of simulation is most commonly used for credentialing purposes due to its suitability for administration to large groups and objectivity in scoring (Assmann, Hixon and Kacmarek, 1979). McGuire, Solomon and Bashook (1976) list the following advantages
of written simulations for competency testing:

(1) perceived relevance - because written simulations correspond more closely to clinical situations, students perceive them as a far more relevant assessment method than conventional multiple-choice examinations.

(2) standardization of the task - with written simulations, it is possible to predetermine the exact tasks and their level of complexity for inclusion in assessment.

(3) broad sampling of performance - because written simulations standardize tasks and focus on the significant aspects of each, it is possible to sample clinical performance on a broad and representative group of clinical problems.

(4) objective rating of performance - because the tasks are precisely defined and preselected, it is possible to develop specific and detailed scoring criteria which can be applied consistently and objectively to all students.

It is important to note that written simulations are not useful for the assessment of simple recall processes; this may be more efficiently accomplished with the use of conventional cognitive knowledge tests (McGuire, 1969; Scanlon, 1978). Rather, written simulations are most useful for assessing higher order cognitive skills such as problem solving ability which is generally considered important to effective health practitioner performance (Andrew, 1972; LaDuca, 1979; Marshall, 1977; Martin, 1975; McGuire, 1976; Scanlon, 1978).

In summary, written clinical simulations may be perceived as realistic work sample tests which can be
used to augment traditional competency assessment methods. They are currently used in addition to multiple-choice type tests by the NBME for physician credentialing in the U.S., by similar national boards in Australia and Canada, and by the National Board for Respiratory Care (NBRC) in the U.S. for credentialing respiratory care practitioners.

**Assessing Clinical Problem Solving Ability**

It has been stated that the quality of clinical judgment rendered by a physician is probably the most important determinant of the quality of care that will be provided (Williamson, 1965). Clinical judgment is synonymously referred to, as well as operationalized for testing, as problem solving ability (Vu, 1980). Sedlack and Nattress (1972) underscored the importance of clinical problem solving ability by claiming that the use of knowledge in solving problems and rendering decisions makes the difference between competent and incompetent performance. Clinical problem solving ability is therefore considered an essential component of competence in physician practice (Andrew, 1972) as well as related health fields such as nursing (DeTornyay, 1968), respiratory therapy (Scanlon, 1978), and pharmacy (Page and Fielding, 1980).

Problem solving is a complex process that is
difficult to observe directly and even more difficult to measure objectively (Scanlon, 1978). Nevertheless, because clinical judgment is felt to be the heart of technical quality (Donabedian, 1976), considerable emphasis has been placed on the teaching and testing of this intangible and somewhat elusive aspect of competence. This is true for physician as well as related clinical practice fields. Certainly the type of clinical problem presented to a practitioner may vary according to health care specialty, but the process of rationally collecting and evaluating data, as well as effectively implementing a decision is ubiquitous in the daily role of a clinical practitioner (Scanlon, 1978). Consequently, as physician training programs and credentialing agencies began emphasizing clinical problem solving ability and using written simulations for teaching and testing this aspect of competence, so too did other health professions.

The first application of pencil and paper problem solving tests to medicine was introduced by Rimoldi (Marshall, 1983), and called the Test of Diagnostic Skills (Rimoldi, 1961). Citing the deficiencies of conventional test methods, Rimoldi (1955) suggested that gaining an understanding of how information is gathered and evaluated by subjects would improve selection and prediction procedures, as well as teaching methods. The
test Rimoldi devised consisted of a card deck with questions a physician might ask regarding a patient's condition on one side of a card and the response to the question on the reverse side (Rimoldi, 1961). The test method allowed for recording the information selected to solve the problem and the order of selection, as well as the final diagnosis rendered (Rimoldi, 1960).

Building on the work of Rimoldi, the NBME developed a new testing method in 1961 to replace the bedside oral previously used for the Part III examination (Hubbard, Levit, Schumacher and Schnabel, 1965). This test method was designed to standardize the content of Part III for all examinees, be objectively scored, and assess the examinee's ability to identify, resolve and manage patient problems. The NBME termed this method "Programmed Testing" as it utilizes a paper and pencil format to simulate a realistic clinical situation (Hubbard et al., 1965). Each patient problem begins with brief but carefully selected information which is followed by lists of possible diagnostic procedures and therapeutic alternatives, some of which are correct and mandatory for management of the patient and others which are incorrect and contraindicated. The task of the examinee is to select as many of the procedures and courses of action as are deemed appropriate for the patient described (Hubbard et al., 1965).
A similar, but more elaborate, written simulation instrument was developed by Christine McGuire and her associates at the University of Illinois in the late sixties (Marshall, 1983). This instrument was called a Patient Management Problem (PMP) and also allowed for assessment of treatment as well as diagnostic skills in medical problem solving, and could be readily administered to large groups of students (McGuire and Babbott, 1967). A PMP attempts to simulate in written form the process a clinician uses in managing a patient (McGuire, Solomon and Bashook, 1976). Each problem begins with a short introduction, called the scenario, which sets the scene for the clinical encounter. This is followed by a series of sections, each of which provides for either gathering information about the patient's condition or selecting among therapeutic alternatives for clinical management. Each section contains a list of options from which the examinee makes selections. Responses to the options were originally concealed by opaque tabs but more recently are hidden by the use of invisible (latent image) ink, and the examinee records his/her selections by removal of the tab or development of the ink with a special pen, thus revealing information requested or results of a therapeutic intervention. In completing the problem, the process is recursive in that the patient situation constantly changes as a result of the
actions taken by the examinee. There are many possible routes through a problem so the order and number of sections completed by each examinee differs according to selections made at divergent points in the problem. Such a problem is termed a branching PMP and allows for multiple approaches to patient management.

McGuire (1976) delineated several specific design characteristics for the PMP instrument developed to assess clinical problem solving skills. First, the introduction to the problem must be presented in a realistic manner which resembles the form in which a patient problem would be posed in an actual clinical situation, and responses to options must include only the results of information requested or of therapy initiated (i.e., not corrective feedback). Also, it must be impossible for an examinee to retract a selection, thus one must deal with the results of clinical decisions in a realistic way such that the patient situation changes as a result of examinee actions. Therefore, the problem must allow for different approaches to patient management, and require a series of sequential, interdependent decisions which may be somewhat unique based on prior selections in the problem.

It is this type of branching PMP that was developed for assessing the clinical problem solving ability of respiratory therapists (American Association of
Respiratory Therapy, 1978), and is used in the NBRC written clinical simulation examination. The branching type of PMP differs considerably from the problem format the NBME uses. For instance, the NBME problems are termed linear PMPs as they do not allow for branching (Barro, 1973), and from an assessment standpoint, focus on the appropriateness of each decision considered independently and on efficiency in reaching a single correct solution (McGuire and Babbott, 1967).

Different scoring methods are also used to describe performance on PMPs. Although the NBME uses 0/1 weighting (i.e., correct/incorrect) for options pertaining to the gathering of data in patient problems (Barro, 1973), commonly PMP options are weighted on a five point scale indicating the relative degree of helpfulness or hindrance to resolution of the patient problem (McGuire, Solomon and Bashook, 1976). Such a scoring system differs considerably from that of traditional objective cognitive knowledge tests and includes negative weights. The positive and negative point values used for the scoring scale vary, but Donnelly (1976) demonstrated that correlations between different weighting systems are very high.

However, the manner in which option weights are tallied results in considerably different summary scores for performance. One method implicitly assumes that
PMPs are unidimensional as single summary performance scores are derived (Donnelly et al., 1974). Although quite a variety of these exist, the two most commonly used were originally proposed by McGuire, Solomon and Bashook (1976). The proficiency score represents the extent to which options selected were helpful in the resolution of the problem; it is determined by summing the weights of all options selected, dividing this sum by the maximum number of positive points possible to obtain in the problem, and multiplying this result by 100 to convert to a percentage score (Assmann, Hixon and Kacmarek, 1979). The other is called an Efficiency score and represents the proportion of positively weighted options that were selected by an examinee; it is determined by counting the number of positively weighted options selected, dividing this value by the total number of options selected (positive and negative) and multiplying by 100 to convert to a percentage (Assmann, Hixon and Kacmarek, 1979). The NBME uses a single summary score for describing performance on PMPs (Hubbard, 1978), although its precise calculation is not reported in the literature.

Another method of deriving summary scores is based on the delineation of the skills sampled in a PMP (McGuire, Solomon and Bashook, 1976). This may be done in a variety of ways but basically the method involves
categorizing sections of the PMP relative to their content, or stage in the care of a patient or problem solving process, and generating subscores which describe performance related to the delineated categories. These scores are derived by simply summing the weights of options selected in sections assigned to the respective categories. The resulting scores may be reported in raw form (i.e., simple total), or the sum divided by the number of options selected in the respective category to obtain an average score. The NBRC reports category scores as simple sums for stages in the problem solving process (i.e., information gathering - IG, and decision making - DM), and for specified content areas respective to IG and DM as well (National Board for Respiratory Care, 1984).

Several varieties of written problem solving exercises are used by training programs for teaching and testing clinical problem solving skills, as well as for research purposes. They differ from linear and branching structured response PMPs in many respects including the focus of the problem solving task, item stimulus and response characteristics and scoring procedures. For instance, the Diagnostic Management Problem (Helfer and Slater, 1971) focuses on diagnostic skills and uses a card deck similar to the Rimoldi method; the Sequential Management Problem (Martin, 1975) is primarily used for
instructional purposes and provides corrective feedback at each stage of the problem (Marshall, 1983). A review by Galofre in 1974 identified eighteen varieties of PMP type instruments (Newble, 1979).

The vast number of, and differences in, written simulation instruments used for research on clinical problem solving ability render it difficult to meaningfully compare results or draw conclusions (Marshall, 1983). Several studies have demonstrated that performance on structured response PMPs is not comparable to performance with other types of simulated patient encounters or actual practice (Vu, 1979). For instance, studies by Martin (1975) and McCarthy (1966) which compared performance on PMPs to simulation formats which required examinees to generate responses rather than select from a list of options, demonstrated a significant cueing effect in that more data was gathered and overall problem solving performance was better in structured response PMPs. This effect was particularly pronounced for students who did poorly on the free response instrument (McCarthy, 1966). This cueing effect was also demonstrated in studies which compared written PMP performance to that observed in physician interactions with actors trained to play patients (Page and Fielding, 1980; Norman and Feightner, 1981), and with performance in actual practice (Goran, Williamson
and Gonnella, 1973). It seems clear that the format of a problem solving exercise affects the behavior of a subject and consequently the nature of what is assessed (Marshall, 1983), and that this is particularly true for written PMPs which offer lists of alternatives from which an examinee chooses (Vu, 1979).

**Credentialing in Respiratory Therapy**

The National Board for Respiratory Care (NBRC) is a private, not-for-profit credentialing agency, which was formerly called the National Board for Respiratory Therapy. According to its by-laws, the NBRC is charged with setting standards by which the competence of respiratory care practitioners may be determined. As sanctioned by the American Medical Association (AMA), and in conjunction with the Joint Review Committee on Respiratory Therapy Education (JRCRTE), an integrated competency assessment system has been developed which provides a definitional basis for competence. The system is hierarchical in that one must progress through each level in a sequential manner. National job analysis studies have identified two levels of clinical practice (Nettles, 1981; 1982), and appropriate measures to assess the relevant components of each have been developed. The structure of this system is presented in Table 1.
The NBRC uses a five step process in the development of a credentialing examination (Lawrence, 1982). First, a viability study is undertaken using a modified nominal group technique to establish the need and purpose for an examination. A manpower study is then conducted to determine the basis for sampling in the job analysis study to follow. The results of the job analysis survey provide the criterion for examination type and content. Once a pilot form of the examination is developed, a criterion-related validity study is conducted in accordance with federal guidelines to establish job-relatedness (U.S. Equal Employment Opportunity Commission et al., 1978). If no revisions are indicated, parallel forms of the test are developed and examinations are offered biannually. Job analysis and criterion-related validity studies are repeated every five years to investigate the currency of the competence criterion, and test revisions are made as indicated.

Graduation from an AMA approved respiratory therapy program is required to qualify for the entry level Certified Respiratory Therapy Technician (CRTT) examination, which consists of 200 basic job knowledge and skill multiple-choice items. Upon successful completion of this examination and one year of documented clinical work experience, a candidate may qualify for the advanced level Registered Respiratory Therapist (RRT)
examination. This consists of two parts, the first designed to measure advanced job knowledge and skills via 100 multiple-choice items, and the second designed to measure clinical problem solving ability in a written simulation format using ten branching PMPs. The components of the nationally standardized competency assessment system for respiratory therapists are summarized in the table below.

### Table 1.

**Respiratory Therapy Competency Assessment System**

<table>
<thead>
<tr>
<th>Source Document</th>
<th>Competency Profile</th>
<th>Assessment Method</th>
<th>Evaluation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>JRCRTE Essentials for an Accredited Educational Program</td>
<td>Medical knowledge and clinical skills</td>
<td>By educational institution (i.e. granting of diploma/certificate of completion)</td>
<td>5 year re-accreditation (i.e. self-study and site visit)</td>
</tr>
<tr>
<td>Entry Level Job Analysis</td>
<td>Entry Level job knowledges and skills</td>
<td>CRTT exam (200 item cognitive knowledge test)</td>
<td>Criterion-related validity study</td>
</tr>
<tr>
<td>Advanced Level Job Analysis</td>
<td>1. Advanced Level job knowledges and skills</td>
<td>RRT written exam (100 item cognitive knowledge test)</td>
<td>Criterion-related validity study</td>
</tr>
<tr>
<td></td>
<td>2. Clinical problem-solving ability</td>
<td>RRT written clinical simulation examination</td>
<td>Criterion-related validity study</td>
</tr>
</tbody>
</table>
The Validity of Written Clinical Simulations

Despite the attempt to delimit competence as problem solving ability, and the widespread use of written simulation instruments, there is considerable controversy regarding the reliability and validity of such measures. Due to the high degree of "face validity" that written simulations possess, it has been commonly asserted that the similarity of clinical problem solving tasks in patient care and on PMPs render such simulations content valid (Swanson, 1984). Similarly, the criterion-related validity of PMP's has been asserted because of the high degree of point-to-point correspondence between the test and the criterion of clinical performance (McGuire, 1976). With the contemporary legal emphasis on job-relatedness in employment testing (U.S. Equal Employment Opportunity Commission et al., 1978), this logic is appealing. However, research in this area has questioned the reliability and validity of written simulation instruments.

Reliability and Generalizability

The traditional concept of reliability refers to the consistency of measurement - the extent to which variation in a set of test scores represents systematic differences among individuals rather than sources of error variation (Stanley, 1971). However, many authors
have suggested that conventional methods of estimating reliability based on classical test theory are not useful for written simulation instruments (McGuire and Babbott, 1967; Sedlacek and Nattress, 1972; Small, 1975; Vu, 1979). McGuire and Babbott (1967) claim that some characteristics of PMPs which act to increase the validity of these instruments for clinical competency assessment, also render conventional reliability estimates inappropriate, such as:

1. items are differentially weighted and interdependent;
2. different options are selected by examinees so the nature of the problem posed varies based on previous selections; and
3. in branching exercises, not all examinees complete the same number or type of items.

Traditional reliability and validity concepts emphasize the independence of each response rather than the complex interrelated responses in a PMP (Sedlacek and Nattress, 1972). For instance, the Kuder-Richardson formulas are not appropriate because the contribution of differentially weighted items to the final score may be negative, zero, or positive; split-half or odd-even correlation methods can not be used because there is no way to identify subtests which would contain equal numbers of items to constitute equivalent forms (Dincher and Stidger, 1976). Additionally, traditional methods of estimating internal consistency assume the
independence of items so this is not entirely applicable (Vu, 1979). Additionally, Donnelly, Gallagher, Hess and Hogan (1974) found that measures derived from a PMP are not unidimensional and suggested it may be more appropriate to determine reliabilities for a given dimension across problems rather than that of a given problem. Bashook (1976) added that estimating reliability for a single problem is not useful because each problem situation presents unique characteristics.

An alternative conception of reliability in terms of viewing the consistency of measurement as the generalizability of scores across problems was suggested by McGuire and Babbott (1967). The importance of generalizability of measures is derived from the notion that the purpose of administering clinical simulations is not merely to assess performance on the particular cases used, but to provide a general assessment of the examinee's ability to solve problems in a defined clinical domain (Swanson, 1984). Therefore, the issue of generalizability is linked to the notion of content validity - that is, how well the content of the test samples the class of situations about which conclusions are to be drawn (Cronbach, 1971). Consequently, the PMPs used to represent a clinical domain must be selected to effectively sample the relevant aspects of a defined domain, in the same way as multiple-choice test
items are viewed as a sample from a universe of items (Swanson, 1984).

However, many studies have demonstrated low inter-case correlations in the range of 0.1 to 0.4 (Swanson, 1984), and this has founded the conclusion that clinical problem solving ability is therefore case specific (Berner, Bligh and Guerin, 1977; Donnelly et al., 1974; Elstein, Shulman and Sprafka, 1978; Marshall, 1977; Skakun, 1978). The notion of case specificity is described by McGaghie (1979) as diagnostic proficiency in medical problem solving depending more on the content of the clinical problem than on characteristics of medical practitioners. Alternatively, there has been considerable criticism of the case specificity conclusion as it is argued that researchers have failed to delineate a definitional basis on which problems vary (Berner and Engel, 1979; LaDuca, 1979; McGaghie, 1979), and therefore groups of problems represent arbitrary aggregates of content and patient situations (LaDuca, 1980).

Moreover, Bashook (1976) has argued that estimating consistency between problems is a useless exercise in test analysis because comparing problems with different content is like comparing items from different subtests to establish the internal consistency of the whole. He claims that attempts to measure clinical problem solving ability have been based on modeling a poorly defined
decision making process rather than defining the domain in which the clinician functions. He argues that the validity of PMPs is really an issue of sampling across the breadth of a domain, and suggests developing a framework for the clinical domain to be assessed which is specific to the respective roles and functions in a medical field and includes the following three dimensions: (1) clinical discipline, (2) context of care, and (3) problem solving process. In this way, the necessary and sufficient conditions for acceptable performance in a domain can be established, and tests devised which representatively sample these elements.

Swanson (1984) has also argued that most simulation based tests are significantly shorter than is psychometrically reasonable (i.e., consist of only a few problems); and many authors have advised that a larger number of problems, drawn representatively from the clinical domain of interest, are required to accurately measure problem solving ability (Barro, 1973; Donnelly et al., 1974; LaDuca, 1979; Norcini, Swanson, Grosso and Webster, 1983; Swanson, 1984). Indeed, McGuire and Babbott (1967) reported improved reliability estimates (i.e., generalizability across tests) by increasing test length: ranges of .75 to .85 for one lengthy problem, .80 to .90 for two and three problem tests designed for a single medical specialty, and .85 to .94 for tests
composed of ten to twelve problems drawn from a variety of medical specialties. The NBME reports reliability estimates (i.e., stability of scores) in the range of .80 to .85 for the PMP portion of their credentialing examination, which is comparable to sections of equal length in the multiple choice parts (Barro, 1973). The Australian physician credentialing agency reported steadily improving intercase correlations ($r = 0.35$ in 1972 ... $0.61$ in 1976) by directing attention to better test development and scoring procedures as well as assuring that difficulty levels across problems were comparable (Marshall, 1977).

Although the PMP reliability and generalizability issue is far from resolved, it appears that closer attention to developing and selecting a sufficient number of representative PMPs to comprise a test may result in more stable scores. As was presented in Chapter One, the NBRC has delineated the clinical practice domain for respiratory therapy, developed a problem pool accordingly, and carefully selects ten PMPs according to strict decision rules to insure content representativeness for each test form. However, because branching PMPs are used and therefore not all examinees complete the same number or sequence of items, conventional reliability estimates for the test as a whole can not be calculated. In an effort to provide some estimate of reliability,
the testing service contracted by NBRC calculates an internal consistency reliability estimate of the IG and DM responses for each route through each problem. A detailed description of the content and reliability estimates of the NBRC Clinical Simulation Examination to be used in this study is reported in Chapter Three.

**Criterion-related Validity Research**

Criterion-related validity studies have contributed further confusion to controversy regarding the validity of PMP tests. A study by Goran, Williamson and Gonnella (1973) compared the performance of 35 emergency room physicians on one PMP to performance in actual practice, using the records of 33 patients seen by the physicians with presenting conditions analogous to that in the PMP. The researchers found that more history and physical data were elicited on the PMP than was recorded on real patients and that the physicians as a group performed consistently better on the PMP. It could be argued however, that time and economic constraints as well as possible incomplete recording by the physicians in the real clinical situation, could have affected these results. Goran et al. (1973) also found that those physicians scoring highest on the PMP were not necessarily the better performers in actual practice. In reaction to this finding, McGuire (1976) argued that such
tests can only predict how one is capable of performing in practice, not how one will perform. Small (1975) elaborated on this by adding that in professional evaluation, we can only deal with probabilities in terms of predicting whether or not a given candidate is likely to be competent in clinical practice, not in certainties.

Alternatively, some criterion-related studies have compared the performance of groups of practitioners on written PMPs to performance on other forms of simulation of the clinical encounter such as short answer examinations or observation of behavior with actors trained to play patients (Blumberg, 1981; Martin, 1975; McCarthy, 1966; Norman and Feightner, 1981; Page and Fielding, 1980). The consistent finding is that subjects perform significantly better on written structured response PMPs than they do on patient problems where it is necessary for them to generate their own alternatives. McCarthy (1966) concluded that the visual cueing in structured response PMPs acts as an aid to recall of relevant information and refers to Ebel's statement that it is reasonable that short answer and choice items yield different kinds of behavior.

In the criterion-related study conducted by Norman and Feightner (1981), which compared the performance of 65 pharmacy students on two written structured response PMPs and two actor/patient stimulus free response
simulations and resulted in a statistically significant increase in options selected/elicited in all sections of the written problems (F test, p< .0001), the researchers also determined the magnitude of the effects of the differences in simulation format, differences across cases and differences between students. The results of the stepwise multiple regression analysis demonstrated that approximately 70% of the variance in options selected/elicited was explained by these three effects; however, differences between students accounted for only 20-30% of the variance in each section of the problem. Overall, the difference in simulation formats accounted for a greater proportion of variance in the total number of options than differences between students or across cases. Moreover, the differences across cases and formats together accounted for 43% of the total variance compared to 26% of the variance resulting from differences between students (Norman and Feightner, 1981).

Marshall (1983) has commented that comparing different forms of simulations implicitly assumes that all measure the same behaviors although it seems clear they do not. Vu (1979) has therefore summarily criticized that the choice of instruments as predictors and their related criteria are inappropriate in such studies.

The NBRC recently completed a criterion-related validation study for the Clinical Simulation Examination
in which the test score(s) served as the predictor(s), and a job performance score derived from a thirty-one item supervisor rating form served as the criterion (Kacmarek, 1985). Statistically significant positive correlations (p< .01) resulted for the total group of subjects (n=308), as well as for sex and race subgroups, and it was concluded that test performance was predictive of job performance in this clinical practice domain (Bryant, 1985). A complete description of this study is included in Chapter Three.

**Research on Construct Validity**

The many measurement difficulties with PMPs have led to an emphasis on construct validity research in an attempt to explicate the meaning of that which is measured by written simulation instruments (Donnelly and Prevot, 1978). Indeed, the investigation of construct validity is particularly important when competence is defined in terms of skills which are hypothesized to underlie specific task performance (Forsythe, 1984). One method of assessing construct validity employs a correlational design in which scores from other measures are hypothesized to be relatively correlated or uncorrelated with scores from the test in question; evidence of convergent validity is assessed by the correlation between performance on the test in question and
performance on other measures of the same trait, and
evidence of divergent validity is assessed by the
correlation between the test in question and measures of
different traits (Campbell and Fiske, 1959). Another
method of assessing the construct validity of a test is
by directly testing implications regarding the abilities
hypothesized to underlie test performance, such as
designs in which subjects are stratified into groups on
the basis of some variable such that differences in test
performance are expected across groups (Cronbach and
Meehl, 1955). Both methods have been used for research
on the construct validity of PMPs.

Small (1975) has stated that clinical effectiveness
and high quality performance are impossible without a
cognitive base, but that the evaluation of clinical
skills stressing diagnosis, patient management and
technical ability must assess more than theoretical
knowledge. Based on the assumption that clinical
problem solving ability is not possible without a back­
ground knowledge (Marshall, 1977; 1983), many studies
have investigated the relationship between PMP perform­
ance and performance on other types of written medical
tests. For instance, Case (1981) demonstrated that the
correlations for a group of fourth year medical students
(n=486) between scores on two written simulations and
total NBME Part I (basic sciences) and Part II (clinical
sciences) multiple-choice test scores, were .30 and .38 respectively. Although described only briefly, McGuire and Babbott (1967) reported relatively low correlations (i.e., generally not significantly different from zero) between scores on PMPs and scores on multiple-choice tests in the same discipline, even when the latter were designed to measure some aspects of problem solving skills. Findings such as these have commonly been interpreted as indirect evidence that PMPs measure something different than traditional multiple-choice type tests, and that PMPs are therefore useful for assessing clinical competence (Swanson, 1984).

In a later article, McGuire (1976) reported in a general manner that studies of performance of medical students, residents and board certification candidates are consistent in revealing low correlations (i.e., .20 to .40) between scores on PMPs and those on multiple-choice tests, even when both sets of exercises are quite lengthy and the multiple-choice items are designed to test similar abilities as PMPs such as interpreting data and solving problems. McGuire (1976) concluded that this is as expected if different aspects of competence are sampled by the two types of tests. Heifer and Slater (1971) further demonstrated that correlations between scores derived from 42 pediatric Diagnostic Management Problems (a variation of the PMP which
utilizes a card deck) were .09 with Pediatric National
Board multiple-choice test scores, .40 with pediatric
clerkship grades (derived from instructor ratings of
clinical performance), and .60 with scores derived from
a pediatric PMP. Thus, an expected divergence was found
between the card deck diagnostic problem solving exer-
cises and a written cognitive knowledge (multiple-choice
type) test, and convergence demonstrated by significant-
ly higher correlations with measures intended to reflect
clinical problem solving skills (i.e., the clinical
performance ratings and the PMP), with all assessments
related to the same clinical practice domain.

On the other hand, a study by Holzemer, Schleuter-
mann, Farrand and Miller (1981) investigated the rela-
tionship between the scores of 79 nurse practitioners on
one PMP and several independent variables including
patient chart audit ratings, self-evaluation ratings,
peer ratings, and a 47 item multiple-choice test whose
content paralleled that of the PMP. The Pearson product
moment and Spearman rank order correlations were
statistically significant only for the relationship
between the multiple-choice test scores and the PMP
scores; \( r = .54 \) and \( \rho = .56 \) respectively. The authors
concluded that these two tests may be measuring the same
construct and recommended the multiple-choice test as a
preferred assessment method because such a test is less
difficult to construct as compared to PMPs. However, a stepwise multiple regression analysis revealed that only 37.6% of total score variance was explained by all the predictors used in the study (Holzemer et al., 1981), perhaps indicating that the use of a score from one PMP may be insufficient to justify such a conclusion.

As a second phase to the above described study, 46 nurse practitioners and 31 registered nurses completed three PMPs, a 70 item multiple-choice test based on the content of the PMPs, and a demographic questionnaire (Farrand, Holzemer and Schleutermann, 1982). The multiple-choice test score was hypothesized to be a meaningful covariate with the PMP scores based on the results of the previous study. However, the multiple-choice test score correlated significantly with only one of the PMPs (r = .49); correlations with the other two PMPs were .13 and .12 respectively. In addition, although expected significant differences in performance were found between the nurse practitioner and registered nurse subgroups (i.e., the nurse practitioner group performed significantly better as expected due to the effect of more education and responsibility in managing patients), there were also unexpected significant differences in performance across the three PMPs within each subgroup which suggested that one measurement characteristic of PMPs is difficulty level (Farrand et al.,
1982), and this may have affected the results of the prior study. A second unexpected finding was a curvilinear relationship between performance on the PMPs and years of experience in the nursing field; subjects with less than two or more than fifteen years of experience performed poorer than the middle group (Farrand et al., 1982). Although the authors did not describe how this finding was demonstrated, it appears that both the nurse practitioner and registered nurse subgroups were grouped together in this part of the analysis, and therefore there may be characteristics of the less than two years experience subgroup and/or greater than fifteen years subgroup which contributed to this finding, such as an artifact of educational level and/or level of responsibility in managing patients.

An earlier study by Williamson (1965) had similarly demonstrated a low negative correlation between years in practice and performance on a PMP in a heterogeneous group of practicing physicians. However, these results are based on performance on one PMP, and not enough information is known about the nature or quality of the clinical experience during the years of practice of the subjects to insure meaningfulness of this finding.

A recently completed dissertation by Kacmarek (1984) investigated the effect of years of clinical experience in respiratory therapy, performance on the
NBRC 200 item multiple-choice test, and some categorical variables such as level of education, region of the country in which educated/employed, etc., on performance on the June 1982 NBRC Clinical Simulation Examination (consisting of ten PMPs). Kacmarek derived a composite score for the two score variables reported by the NBRC to describe performance on the PMP test as a dependent variable, and data on twelve independent variables were obtained including those described above for more than 2000 candidates. Results of a multiple regression analysis and discriminant analysis revealed that performance on the multiple-choice examination was most predictive of PMP performance and best able to classify subjects into pass/fail groups; also, increasing years of experience prior to taking the Clinical Simulation Examination was negatively correlated to performance on this PMP test (Kacmarek, 1984). However, the zero order correlation was .28 between the multiple-choice test score and the PMP composite score, accounting for only 8% of the total PMP score variance; a correlation of .37 resulted between the multiple-choice test score and years of experience variables considered together and PMP performance, which accounts for approximately 14% of the total variance (Kacmarek, 1984). Despite the magnitude of these correlations, a small proportion of variance is explained by these two variables even in combination,
and it is possible that subjects with increasing years of experience prior to taking the PMP test did so due to a self-recognized weakness in clinical ability.

Although the issue of the role of clinical experience and the relationship between scores derived from traditional objective cognitive knowledge tests and performance on PMPs remains somewhat unclear, Marshall, Fleming, Heffernan and Kasch (1982) argue that the various tests used for research in this area vary widely in content, format and scoring procedures, possibly influencing outcomes. Marshall et al. (1982) designed a study utilizing two PMPs and two true-false examinations reflecting the content of each PMP directly, which were developed with the assistance of content experts specifically for the study and were pilot tested on a separate group of experienced physicians prior to the study. The researchers selected two groups of subjects appropriate to the purposes of the study; one group consisted of eight residents in family practice and the other of six experienced general practitioners. Each subject completed the two PMPs first, followed by completion of the true-false tests. The results indicated that both groups performed essentially the same on the true-false tests, but that the experienced physician group performed significantly better on both PMPs (Marshall, et al., 1982).
Several other studies have investigated the effect of experience on PMP performance by using stratified samples of known experience as opposed to relying on years of experience as an independent variable. Very early on in the experimentation with simulated problem solving exercises, Rimoldi (1955; 1961) demonstrated that groups with known experience in the substantive domain of the problem to be solved were more discriminate in eliciting data upon which to base diagnoses. McGuire (1976) also reported that groups with more clinical education and experience tended to require less diagnostic information and were more willing to take decisive action in treatment.

The NBME undertook a cross-sectional and longitudinal investigation to determine if increasing amounts of clinical experience would affect PMP performance in a predictable way (Hubbard, 1978). It was hypothesized that examinees near the end of their internship year should perform better on the NBME Part III PMP test than those near the end of their third year of medical school, as the former group would have experienced more responsibility in managing patients. The results of the cross-sectional analysis indicated that 90% of the third year student's scores fell below the mean of the internship group (Hubbard, 1978). The third year student group was retested on a parallel Part
III PMP test form near the end of their internship year, and this longitudinal analysis revealed that 91% of the scores achieved by the third year students fell below the mean of the scores of this group retested in their internship (Hubbard, 1978).

Of interest in this area is a study conducted by Alderman, Evans and Wilder (1981) which investigated similar relationships regarding the performance of law students on written structured response simulations of client encounters. Six simulations were designed for the study, some involving civil cases and others minor criminal offenses. A large stratified sample of subjects consisting of 202 pre-law undergraduate students (assumed to have no client experience), 157 law students without client experience and 255 law students with client experience, completed the simulations. Although there was no significant correlation between scores achieved on the simulations and current grade point average for the experienced group, this group did perform significantly better than both inexperienced groups and it was determined that the nature of the client experience of the subjects in the experienced group (i.e., civil versus criminal) resulted in better scores on the problems most related to the nature of one's experience (Alderman et al., 1981).

The many differences in PMP format, content and
scoring procedures of the instruments used in these studies, as well as differences in the type and nature of independent variables, result in a fragmented body of research upon which it is difficult to render generalizable conclusions. Indeed, construct validity is not established by confirming a single prediction on different occasions or many predictions in a single study; rather evidence of construct validity is demonstrated by a pattern of consistent findings across a number of different studies (Carmines and Zeller, 1979). Although it seems reasonable to conclude that medical knowledge and clinical experience contribute to clinical problem solving ability to an extent, the precise nature of this relationship remains elusive and research results appear to be affected by the choice of independent variables and the possible unreliability of variables used in many studies. Vu (1979) offers that research in this area has suffered from the lack of a commonly agreed upon definition of effective problem solving ability, and subsequent variation in its measurement. Vu (1979) concludes that the development of effective methods to assess problem solving ability depends on the capability of researchers to both theoretically and empirically define effective problem solving behavior. Unfortunately, there is considerable controversy based on research regarding theories of problem solving as well.
Theories of Human Problem Solving

Underlying the effort to measure clinical problem solving ability with written simulation instruments is an assumption that there is a common problem solving process which can be applied to different clinical content (Berner, Bligh and Guerin, 1977). However, a great deal of confusion currently exists regarding the extent to which medical knowledge and problem solving ability underlie the solution of clinical problems (Harasym, Baumber, Bryant, Fundytus, Preshaw, Watanabe and Wyse, 1980). The results of research on medical problem solving ability generally fall into one of two categories: (1) clinical problem solving is primarily dependent upon medical content specific within a case, or (2) clinical problem solving is a skill, or series of skills, which can be applied to all problems (Harasym et al., 1980). As problem solving ability can not be dependably taught or evaluated without a clear understanding of this phenomenon (Elstein, Shulman and Sprafka, 1978), research on the cognitive nature of problem solving is central to refining techniques for its assessment. The purpose of this section of the review is to present the research upon which opposing theories of medical problem have been founded; relevant and related research in psychology is presented first to aid in conceptual clarity.
Research in Psychology

A paper written by Newell, Shaw and Simon (1958) delineated the questions that a theory of problem solving should answer. Among these were that it should: (1) predict the performance of a problem solver handling specified tasks, (2) explain how human problem solving occurs including the processes that are used and the mechanisms that perform these processes, and (3) show how changes in attendant conditions, both intellectually within the subject and externally regarding the task stimulus, alter problem solving behavior (Newell, Shaw and Simon, 1958).

In a later article, Simon and Newell (1971) reviewed the research which attempted to answer these questions and thereby presented the state of the information processing theory of human problem solving in 1970. The central emphasis of this research was on the process of how particular human behaviors occur and the mechanisms which enable them. Simon and Newell (1971) characterize the information processing theory of human problem solving as consisting of three major components: (1) the information processing system (i.e., short and long term memory), (2) the task as defined in terms of a task environment, and (3) the problem space as defined intellectually by the problem solver.

The four major propositions of the theory are that:
(1) only a very few gross characteristics of the human information processing system are invariant over tasks and problem solvers, (2) the problem solver represents the task environment as a problem space, (3) the structure of the task environment determines the possible structure of the problem space, and (4) the problem space structure determines the information (i.e., memory programs) that will be used for problem solving. Also noted was the importance of experience with the type of problem to be solved in constructing problem spaces. Moreover, the system is described as adaptive in that problem solving behavior is determined by the demands of the task environment rather than by some internal human characteristic, and therefore the structure of the task posed to the problem solver is considered as important in the study of problem solving behavior as the process the subject uses.

Hammond, Hursch and Todd (1964) had earlier demonstrated, in studying the clinical judgement of psychologists using tests followed by verbal introspection, that the structure and the dynamics of the clinical task posed affected performance. They concluded that both the test and the clinician are interacting parts of the same system, and therefore judgement was a function of the relationship between the characteristics of the test situation and characteristics of the response system of
the clinician. In a similar manner, other authors refer to the person-environment interactive paradigm in studying individual differences, wherein behavior is perceived as a function of both the person and the environment (Hunt, 1975; Mitchell, 1969).

Hammond, Hursch and Todd (1964) also reported that they found progressive differences in the performance of the clinical psychologists with increasing experience; that is, the more inexperienced clinicians tended to use a systematic textbook rule style in rendering clinical judgements, while the experienced clinicians used an inductive process based on rules generated from clinical experience. DeGroot (1965), following a study of the behavior of chess players, similarly concluded that experience was clearly related to expertise in solving chess problems. Elstein, Kagan, Shulman, Jason and Loupe (1972) noted that science, chess and medicine are all similar in that decisions must necessarily be based on incomplete evidence and therefore, consistent findings regarding problem solving behavior in medicine would not be surprising, as results of the Medical Inquiry Project subsequently demonstrated.

The Medical Inquiry Project

An extensive five year program of research on medical problem solving was conducted under the direction of
Arthur Elstein at Michigan State University called the Medical Inquiry Project (Elstein, Shulman and Sprafka, 1978). Elstein conceived the medical work-up as a form of inquiry with problem solving strategies employed by physicians to reduce uncertainty. Following completion of this program of research, it was concluded that information and experience were basic components of clinical problem solving, as possession of relevant bodies of information and sufficiently broad experience with related clinical problems appeared to affect ability (Elstein et al., 1978). Similar to DeGroot's (1965) conclusion following his study of chess players, Elstein et al. (1978) concluded that "the differences between experts and weaker problem solvers are more to be found in the repertory of their experiences, organized in long-term memory, than in differences in the planning and problem-solving heuristics employed" (p. 276).

One major portion of the Medical Inquiry Project consisted of an indepth descriptive analysis of the reasoning process of 24 experienced physicians on a few medical problems (i.e., observation of physician interactions with actors trained to play patients followed by stimulated recall sessions). Based on this analysis, it was concluded that the sequence and type of strategies employed to solve medical problems varied according to the nature of the situation posed; however, a general
approach to problem solving was characterized as a four stage process consisting of the following:

(1) cue acquisition - the observable component of clinical reasoning in which clues, clinical findings and data are obtained.

(2) hypothesis generation - physicians begin to generate tentative hypotheses in the earliest moments of the interaction with patients, which serve to define the problem space of possible solutions.

(3) cue interpretation - data is interpreted in light of the hypotheses as the work-up proceeds; each hypothesis implies a list of probable features and data obtained are evaluated in terms of correspondence to or departure from the specifications of these lists; some information is simply ignored or dismissed as noncontributory.

(4) hypothesis evaluation - data subsequently collected are thus used to evaluate the hypotheses and if necessary, to generate/formulate new ones, thereby modifying the problem space until a diagnosis is reached.

This is referred to as the hypothetico-deductive model of medical problem solving, and the most striking feature is that of early hypothesis generation (Elstein et al., 1978). As Elstein, Shulman and Sprafka (1978) explain, the medical education model for teaching problem solving implicitly assumes that the competent physician reserves judgement until all pertinent information is gathered and that problem solving is a process common to all clinical cases. The findings of this research contradict these assumptions, and illustrate congruence with the information processing
theory of human problem solving. For instance, before data collection is far advanced, hypotheses are generated which serve to delimit the problem space and guide further inquiry; data gathering is then directed toward confirming or eliminating hypotheses, thereby further delimiting the problem space. Also, the type of clinical problem posed affected both the route and nature of the end point in medical problem solving in lieu of a problem solver preference or cognitive style (Elstein et al., 1978).

Another portion of the Medical Inquiry Project analyzed the performance of 15 of the original 24 physicians on four written PMPs. The consistent finding with these analyses was that individual physician performance varied greatly from case to case. This high degree of intra-individual variation led the researchers to believe that problem solving may be case, and therefore medical content, specific. Moreover, it was generally concluded at the completion of the project that the low intra-individual consistency across problems with both the actor based and written simulations indicated that the problem solver's representation of the task environment strongly determined how problems were approached and that competence in solving medical problems is specifically case related; therefore it was summarily concluded that medical problem solving ability can not
be generalized beyond the cases presented in any given test (Elstein et al., 1978).

Although the Medical Inquiry Project is the only extensive long term study in this area and clearly contributed to a better understanding of medical problem solving, many authors have criticized the finding of case specificity as a methodological artifact (Berner and Engel, 1979; Harasym et al., 1980; LaDuca, 1979, 1980; Marshall et al., 1982; McGaghie, 1979; Vu, 1979, 1980). For instance, Harasym et al. (1979) suggested that the finding of case specificity could be a product of the educations of the physicians investigated because the research focused on indepth analyses of small groups of physicians from the same locale. Berner and Engel (1979) also criticized the researchers sampling of patient problems as they failed to formulate a definitional basis (i.e., universe of patient situations) for observation. Although LaDuca (1980) concedes that the simulations were relevant, he claims they lacked content validity and refers to a double sampling problem (i.e., inadequate subject and clinical problem sampling). Indeed, Elstein et al. (1978) reported that "our problems were chosen with little reference to their content, since it was assumed initially that a common problem-solving approach could be detected in the solution of any medical problem" (p. 117).
Moreover, LaDuca (1980) further argues that although careful sampling is essential to sound methodology and the clinical problems used can be considered no more than an arbitrary aggregate, he contends that the problem with the case specificity conclusion was not simply methodological, but also paradigmatic. That is, because the focus of the study was on an assumed underlying cognitive trait, the design of the research did not allow for an investigation of the relationship between the subject and the clinical problem. LaDuca (1980) advises that the results of the Medical Inquiry Project render it difficult to retain a belief in medical problem solving as a cognitive factor, and draws attention to the notion that human problem solving is a person-environment interaction which requires careful sampling in both directions (i.e., subjects and problem tasks) in order to meaningfully research. This is consistent with the findings of Simon and Newell (1971), as well as others discussed in the previous section, regarding the nature of human problem solving in psychology and the importance of the task environment.

**Process Theories of Medical Problem Solving**

Several studies have used factor analysis as a "discovery procedure" to investigate the underlying dimensionality of measures derived from medical problem
solving exercises (Swanson, 1984). These studies have used larger subject and patient situation samples, resulting in theories which oppose the notion of case specificity although the design of these studies did not provide for indepth evaluations regarding the person-environment paradigm. Additionally, each study used different subjects, instruments, derived measures and analytic procedures, and results have varied.

The first study of this type was reported by Donnelly, Gallagher, Hess and Hogan (1974). The researchers considered that determining the dimensionality of scores derived from problem solving exercises was a pre-evaluation problem since it was inadvisable to evaluate problem solving ability without an understanding of the phenomenon to be measured and the measurement properties of instruments.

In this study, ten branching PMPs were administered to 162 third year medical students. The various sections of each problem were categorized as pertaining to one of seven stages in the work-up and treatment of a patient, and a score was derived for each student in each category, both for the individual problems and an average across all problems. Scores were derived by summing the item weights which ranged from -2 to +2 as determined by a panel of judges. Separate factor analyses were then performed on the resulting 7x7
correlation matrices - one analysis for each of the individual problems and one for the average scores across problems. For the average score analysis, as well as eight of the ten individual problem analyses, a two factor structure was identified. Factor I was termed information gathering and consisted of the history, physical examination, laboratory and radiographic section scores; factor two consisted of the diagnosis, pathway and management scores and was called decision making (Donnelly et al., 1974). As a conservative test of the orthogonality of the factors, canonical correlations were obtained between the variables comprising each factor both for the average scores across all problems and for the eight problems which displayed a clear two factor structure. Only two of these correlations were significant; the insignificant correlations ranged from .17 to .33. These results founded the theory that two scores are required to adequately describe problem solving ability as assessed by performance on PMPs, and refuted the findings of case specificity as inappropriate because single summary scores had been used which assumed unidimensionality.

A later study by Juul, Noe and Nerenberg (1979) confirmed a two factor structure and further demonstrated it to be stable over time (i.e., one group of medical students took two forms of the examination one
year apart), and across samples (i.e., two groups took the same examination). One examination consisted of 24 branching PMPs and was completed by 191 junior medical students; the second examination consisted of 26 branching PMPs and was completed by this same group of juniors one year later when they were seniors, as well as another group of 214 junior year students.

Total scores across all PMPs were derived for six categories similar to the Donnelly et al. (1974) method (laboratory and radiographic studies were collapsed into a single category called diagnostic studies), and item weights were assigned a value ranging from -8 to +8 based on faculty judgement. Iterative principal factor analyses, using squared multiple correlations of each variable with the other five variables as initial communality estimates, were performed separately on the resulting three 6x6 correlation matrices. For each of the three analyses, two factors emerged that jointly accounted for 71-72% of the total variance, but the factors in all analyses were correlated in the range of .42 to .53. Similar to the Donnelly et al. (1974) results, the history and physical score variables loaded on one factor and the pathway, treatment and diagnosis score variables loaded on the second factor; however, the diagnostic procedures score variable loaded on both. Nevertheless, the researchers concluded that their
results indicated that there are two components to medical problem solving ability as measured by PMPs (i.e., skills in data gathering and skills in patient management), and that at least two subscores are necessary to describe performance as the findings of Donnelly et al. (1974) had suggested (Juul et al., 1979).

Skakun (1978) performed two separate principal components analyses on scores derived from five linear PMPs. Items were weighted on a five point scale (-2 to +2) and four content area scores were determined (history, physical, laboratory and management) for each problem (total of 20 scores), as were four composite scores across all problems. Components with eigenvalues greater than one were retained and transformed by the Harris-Kaiser procedure for each of the two data sets. The results of the analysis on the four scores across all problems demonstrated a two factor solution accounting for 63% variance which was substantively consistent with that of Donnelly et al. (1974) and Juul et al. (1979).

Results of the analysis of the 20 content area individual problem scores did not result in an interpretable structure. It should be noted, however, that each of the five PMPs represented a different medical specialty so perhaps the modest intraindividual consistency noted across problems should be expected. Indeed, Donnelly
and Prevot (1978) demonstrated that predictable and consistent differences in performance related to the stages of the patient work-up and management could be found based on the context of care presented in PMPs (in this case, emergency versus non-emergency situations). They therefore advised, as have many others, that PMP tests must contain a large number and wide variety of representative problems from the domain of interest in order to obtain a meaningful indication of ability.

Further evidence for a process theory of medical problem solving is provided by two other factor analysis studies (Berner, Bligh and Guerin, 1977; Harasym et al., 1980), although free response patient problem solving exercises were used and a four factor structure resulted. These studies were based on the responses of groups of medical students to a series of cases drawn from the same medical practice domain. Results of both the Berner, Bligh and Guerin (1977) and the Harasym et al. (1979) studies demonstrated a four factor structure in which tasks associated with formulating a problem list loaded onto factor I, those involved with selecting correct lab procedures loaded on factor II, avoidance of incorrect lab procedures constituted factor III, and factor IV was exemplary of rendering a final diagnosis. Thus, the notion of case specificity in medical problem solving was refuted by the conclusion that similar
clinical tasks tend to cluster together irrespective of the patient problem, indicating a process dimension to this skill (Berner et al., 1977; Harasym et al., 1979).

Chapter Summary

Due to the difficulties with applying traditional concepts of reliability and validity to PMPs which differ considerably from traditional types of cognitive knowledge tests (Small, 1975; McGuire and Babbott, 1967), and to the many different instruments, scoring methods and analytic procedures used in the great diversity of research on written simulation instruments designed to measure clinical problem solving ability (Donnelly et al., 1982; Swanson, 1984), it is difficult to draw meaningful conclusions about the validity of these instruments as a whole.

The external validity (generalizability) of most of these studies is suspect due to what LaDuca (1980) has referred to as the double sampling error problem; that is, limitations in both sampling of subjects and sampling of observations. Most studies have either used small, rather poorly defined subject samples or large homogeneous groups of individuals readily accessible at a medical school. Additionally, many studies have used only one or a few problems which can not be regarded as
representative of the clinician's domain of practice. Even when larger numbers of problems have been used, little information is provided to enable one to determine the characteristics or degree of representativeness of the test aggregate. Consequently, many variables within and across studies confound the interpretation of research results, both singularly for a given study and collectively for research in this area.

Vu (1979) argues that a systematic approach in considering the different properties of each instrument is necessary and summarizes the situation as follows:

In conclusion, available data on the instruments of each method are incomplete, as is evaluation. In order to obtain a complete evaluation of each instrument, future research will need to provide data on validity or reliability, scoring procedures, the evidence of practicality (for example, time to develop and score, amount of required manpower) and, finally, the level of students for whom an instrument is appropriate. As the task of choosing an appropriate instrument is far more complex than it seems, such complete data along with a careful and systematic approach in considering the different properties of each instrument will be essential to insure the appropriate decision. (p. 305)

Additionally, it has been criticized that most of the research in this area has been specific to the physician practice domain and focused on diagnostic skills (Burke and Connelly, 1981). Indeed, the literature reveals little research directed at understanding the nature of clinical problem solving outside of the domain
of physician practice. The existing models of medical problem solving ability have generally attempted to define one or more of the following: (1) processes that compose clinical problem solving, (2) skills that constitute the processes, and/or (3) variables that affect both (Vu, 1980). The processes involved in clinical problem solving ability have been studied in a fragmented manner, and although none of the models have been investigated in their entirety (Vu, 1980), it appears that an individual's problem solving ability is affected by both the nature of the problems to be solved and experience in the problem domain.

LaDuca (1980) has stated that the organization of behavior, "especially that which constitutes professional competence, is to be found neither totally in the person or totally in the situation ... it is, instead, a relationship between the two" (p. 285). Bashook (1976) has similarly argued that observed variability in performance may be due to the different domains of clinical problem solving, and that the appropriate approach to investigating this ability is to define the domain in which a clinician functions, rather than modeling an undefined process which is presumed to govern it.

Page and Fielding (1980) commented that the controversial evidence regarding the nature of clinical problem solving ability and the validity of PMPs may be
a problem of the state of the art in assessing this component of competence, and Marshall (1983) advised that more research is needed in this area before PMPs are discarded for what may be the wrong reasons. Therefore, it is the purpose of this study to investigate the psychometric properties of a standardized written simulation examination, consisting of ten branching PMPs, which has been systematically developed to assess the clinical problem solving ability of practitioners in the domain of respiratory therapy, and to determine the relationship of quantified measures of relevant work experience to performance on this test.
III. METHODOLOGY

The purpose of this chapter is to present the design, instrumentation and analysis procedures for the study. As the study consisted of two distinct parts, an overview of the design will be presented first to insure clarity. The design section presents the data sources and objectives of the study, as well as descriptions of the population and samples. The remaining sections of this chapter describe the instruments used in the study and the data analysis procedures plan.

Design of the Study

The study was designed to investigate selected psychometric properties of the NBRC Clinical Simulation Examination (CSE). Although NBRC examinations are developed, administered and scored with technical assistance from ETS, the examinations are solely owned by the NBRC. For test security reasons, the NBRC will not release any test forms to independent researchers. Consequently, only that test data made available to the researcher by the NBRC could be used for this study.
**Data Sources**

The NBRC agreed to release test data, without individual identification, for all candidates who completed the CSE in June 1984 for credentialing purposes. This examination was administered to approximately 2800 candidates under standardized conditions at seventy test center locations throughout the U.S.

In exchange for the researcher voluntarily completing a confidential criterion-related validity study of the CSE in accordance with the "Uniform Guidelines on Employee Selection Procedures" (U.S. Equal Employment Opportunity Commission et al., 1978), the NBRC further agreed to allow the researcher to use test data from that study for dissertation purposes. A study group of approximately 300 practitioners completed the June 1984 form of the CSE under standardized test conditions in twelve hospitals during July and August of 1984. The results of the criterion-related study are presented later in this chapter. This administration of the June 1984 CSE form was solely for research purposes; no credentialing decisions were rendered based on results.

The NBRC would not release the names or addresses of the candidate group who completed the June 1984 CSE as this is prohibited by corporation bylaws. However, the NBRC agreed to provide each subject who participated in the criterion-related validity study the Work
Experience Inventory (WEI) instrument for completion and return by mail. In this way, the researcher obtained CSE and WEI data with the ability to match CSE and WEI scores for each respondent of the study subjects.

The NBRC provided a WEI to each of the study subjects in early January 1985, and completed instruments were returned via mail by the end of March. The scoring of all CSE tests and WEI instruments was completed by ETS according to standard procedures. Computer tapes of the results of the NBRC candidate group on the CSE, and the study group on the CSE and WEI, were forwarded to the researcher as they became available. Also, it was determined that fifteen subjects were common to both testing groups; that is, fifteen candidates who took the CSE in June 1984 for credentialing purposes also participated in the study and completed the same test form six to eight weeks later. Therefore, the CSE scores obtained by these individuals on the second test administration and results of their WEI (if completed) were ignored, thereby eliminating these subjects from the study group.

**Population and Samples**

The target population for the study consisted of all respiratory care practitioners actively engaged in the delivery of respiratory therapy to patients in
hospitals in the United States. For the reasons described above, two samples were used in the study. The group of some 2800 individuals who took the June 1984 CSE for credentialing purposes are termed the "candidate" sample; the group of approximately 300 practitioners who completed the June 1984 CSE form for the criterion-related validity study and were subsequently provided a WEI to complete, are termed the "study" sample. Additionally, that subset of the study sample which returned a completed WEI are referred to as the "respondent group".

Representativeness of the Candidate Sample. No information was available to the researcher regarding the biographic or demographic characteristics of the 2821 subjects in the candidate sample because the NBRC does not collect information such as sex, age or race for candidates applying to an examination. Also, for reasons previously explained, the NBRC would not release the names or addresses of candidates so sex or geographic representativeness could not be determined. However, this sample is comparable to previous candidate groups who have seated the CSE in the last few years in terms of the total number of candidates, proportions of first time and repeater candidates seating the test, and pass/fail percentages. Additionally, all CSE candidates must meet pre-established admission criteria including
successful completion of the entry-level Certified Respiratory Therapy Technician (CRTT) examination, documentation of educational requirements and one year of work experience under licensed medical supervision. Therefore, the candidate sample may be considered a somewhat homogeneous group and representative of that subset of the population of practitioners who meet these examination admission requirements.

Representativeness of the Study Sample. Two stage cluster sampling was used to select this sample for the criterion-related validity study. Each of the nine American Hospital Association (AHA) regions of the U.S. was considered a cluster of hospitals, and each hospital within an AHA region considered a cluster of respiratory care practitioners. Due to the concern for test security, the NBRC required that a past or present NBRC Board member be either employed by or professionally associated with, each hospital considered for selection; such individuals are generally managers or medical directors of respiratory care departments. The hospitals were selected so as to be representative of respiratory care practice in the U.S. in terms of geographic location, bed size and institution type, and respiratory care employee demographics. One or two hospitals were selected from each AHA region in order to provide for approximately 30-50 subjects per region. A total of 312 study
subjects completed the experimental administration of the CSE at their place of employment in late July or early August of 1984. Table 2 below provides a description of the hospitals selected for the study.

Table 2. Description of Participating Hospitals

<table>
<thead>
<tr>
<th>AHA Reg/Name</th>
<th>Location</th>
<th>Type</th>
<th>#Beds</th>
<th>#Subj</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/Beth Israel Hospital</td>
<td>Boston MA</td>
<td>Private/Teaching</td>
<td>452</td>
<td>28</td>
</tr>
<tr>
<td>2/Baltimore City</td>
<td>Baltimore MD</td>
<td>Community/Teaching</td>
<td>358</td>
<td>6</td>
</tr>
<tr>
<td>2/Providence Hospital</td>
<td>Columbia SC</td>
<td>Private/Non-Teaching</td>
<td>239</td>
<td>18</td>
</tr>
<tr>
<td>3/Rochester General</td>
<td>Rochester NY</td>
<td>Community/Teaching</td>
<td>500</td>
<td>25</td>
</tr>
<tr>
<td>4/Northwestern University</td>
<td>Chicago IL</td>
<td>University/Teaching</td>
<td>1000</td>
<td>48</td>
</tr>
<tr>
<td>5/University Hospital</td>
<td>Jackson MS</td>
<td>University/Teaching</td>
<td>538</td>
<td>42</td>
</tr>
<tr>
<td>6/St. Luke's Hospital</td>
<td>Kansas City MO</td>
<td>Private/Non-teaching</td>
<td>674</td>
<td>37</td>
</tr>
<tr>
<td>7/St. Luke's Hospital</td>
<td>Houston TX</td>
<td>Private/Teaching</td>
<td>1300</td>
<td>31</td>
</tr>
<tr>
<td>8/Desert Springs</td>
<td>Las Vegas NV</td>
<td>Private/Teaching</td>
<td>220</td>
<td>10</td>
</tr>
<tr>
<td>8/Mercy Medical</td>
<td>Denver CO</td>
<td>Private/Non-teaching</td>
<td>386</td>
<td>22</td>
</tr>
<tr>
<td>9/San Bernardino</td>
<td>San Bernadino CA</td>
<td>Community/Non-teaching</td>
<td>322</td>
<td>21</td>
</tr>
<tr>
<td>9/VA Medical Center</td>
<td>Loma Linda CA</td>
<td>Federal/Teaching</td>
<td>494</td>
<td>20</td>
</tr>
</tbody>
</table>
As all employees of the respiratory care department in a selected hospital were invited to participate, the subjects must be considered volunteers. The study sample consisted of practitioners with quite diverse educational and work experience backgrounds, and was a more heterogeneous group of subjects than the candidate sample. That is to say, although 66% of the study sample had either already achieved the Registered Respiratory Therapist (RRT) credential or were eligible for the CSE, 33% of the study subjects did not meet the admission requirements for this examination (i.e., did not possess the required education and/or experience). Due to the proportion of subjects in the study sample not eligible for the CSE, this sample is considered representative of the respiratory therapy labor force in the U.S., and it was anticipated that the study sample would not perform at as high an ability level as the candidate sample.

Each of the 312 subjects in the study sample were provided a WEI instrument to complete and return via mail. However, only 116 subjects returned the instrument despite two follow-up calls at one month intervals to the hospital contact person. Of these respondents, four had to be eliminated for future analysis; two because they failed to complete entire sections of the instrument, and two who had been previously excluded because they were common to both the candidate and study
samples. Therefore, there were 112 respondents with usable WEI data resulting in a return rate of 35.9%.
The entire study sample had a range of 6 months to 11 years of work experience in respiratory therapy and a mean of 6.61 years; the subset of the study sample which returned a usable WEI (termed the "respondent" group) had the same range of years experience but a slightly lower mean of 5.7 years.

Table 3 provides a description of the entire study sample and that subset of respondents who returned a usable WEI. As can be seen from the table, the proportions of the respondent group as compared to the entire study sample remained essentially the same for sex and very similar for age. The proportions for race and job title are somewhat similar, although the minority and technician subgroups are noted to be less well represented in the respondent group, and supervisors more highly represented in this group.

Additionally, the proportions are noted to be similar for NBRC status except for the "others" category for the respondent group; it may be that the WEI did not apply well to these individuals resulting in less motivation to complete it. Representation by AHA region differed also, particularly for regions 5 and 8 which were under represented in the respondent group and regions 3 and 9 which were over represented.
<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Percent of Study Sample (n=312)</th>
<th>Percent of WEI Respondent Group (n=112)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AHA Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.0</td>
<td>12.5</td>
</tr>
<tr>
<td>2</td>
<td>7.7</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td>8.0</td>
<td>15.2</td>
</tr>
<tr>
<td>4</td>
<td>16.0</td>
<td>12.5</td>
</tr>
<tr>
<td>5</td>
<td>13.5</td>
<td>3.6</td>
</tr>
<tr>
<td>6</td>
<td>11.9</td>
<td>10.7</td>
</tr>
<tr>
<td>7</td>
<td>9.9</td>
<td>15.2</td>
</tr>
<tr>
<td>8</td>
<td>10.3</td>
<td>0.9</td>
</tr>
<tr>
<td>9</td>
<td>13.7</td>
<td>25.0</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>61.0</td>
<td>63.4</td>
</tr>
<tr>
<td>Male</td>
<td>39.0</td>
<td>36.6</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority</td>
<td>31.2</td>
<td>17.9</td>
</tr>
<tr>
<td>Non-minority</td>
<td>69.8</td>
<td>82.1</td>
</tr>
<tr>
<td><strong>Age in Years</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-26</td>
<td>31.5</td>
<td>27.7</td>
</tr>
<tr>
<td>27-34</td>
<td>44.8</td>
<td>53.6</td>
</tr>
<tr>
<td>35-40</td>
<td>15.4</td>
<td>11.6</td>
</tr>
<tr>
<td>Over 40</td>
<td>7.4</td>
<td>7.1</td>
</tr>
<tr>
<td><strong>Job Title</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technician</td>
<td>20.0</td>
<td>9.8</td>
</tr>
<tr>
<td>Therapist</td>
<td>50.0</td>
<td>43.8</td>
</tr>
<tr>
<td>Supervisor</td>
<td>19.0</td>
<td>30.4</td>
</tr>
<tr>
<td>Other</td>
<td>11.0</td>
<td>16.0</td>
</tr>
<tr>
<td><strong>NBRC Status</strong>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRT</td>
<td>40.4</td>
<td>53.6</td>
</tr>
<tr>
<td>CRTT</td>
<td>63.0</td>
<td>68.8</td>
</tr>
<tr>
<td>RRT eligible</td>
<td>25.6</td>
<td>25.9</td>
</tr>
<tr>
<td>CRTT eligible</td>
<td>14.2</td>
<td>10.7</td>
</tr>
<tr>
<td>Other</td>
<td>10.2</td>
<td>0.9</td>
</tr>
</tbody>
</table>

* Note: More than one category may apply.
Objectives of the Study

As the reader will recall, two performance scores are derived for each of the ten problems that comprise the CSE - an information gathering (IG) and decision making (DM) score. This results in a total of twenty problem scores which may be further reduced to single IG and DM summary scores by summing the ten problem scores within each category. The first part of this study investigated the psychometric properties of the twenty problem scores (dependent variables) by: (a) testing for group comparability in terms of the ability level demonstrated by each of the samples respectively on the test, (b) conducting factor analyses in order to determine the dimensionality of measures derived from two administrations of the same form of the test to different groups respectively, and (c) determining reliability (internal consistency) estimates for clusters of related score variables for each sample and each factor.

This design allowed for investigation of the ability level, structure and stability of factors, and reliability of measures derived from the CSE across groups. Consequently, part one of the study used CSE data from both the candidate and study samples. Objectives for this part of the study were as follows:

1. To determine the comparability of groups in terms of ability level by testing for significant differences in test score variance between samples.
2. To determine the dimensionality of the twenty problem score variables derived from an administration of the CSE to a large and relatively homogeneous group of NBRC candidates.

3. To determine the dimensionality of the twenty problem score variables derived from an administration of the same form of the CSE to a smaller more heterogeneous group of research subjects.

4. To determine the internal consistency of the score variables comprising each resulting factor, respectively for both groups.

The second part of the study was designed to investigate the relationship between work experience and test performance. The results of part one provided the basis for determining appropriate summary measures to describe CSE performance for use in part two. For example, if results of the factor analyses demonstrated a stable two factor structure consistent with the hypothesized two process (IG and DM) model, the total test IG and DM scores would be used. On the other hand, if an alternative factor structure resulted from the part one analyses, appropriate summary scores were to be derived based on the problem score variables which comprise each factor. Therefore, based on the results of the factor analyses, appropriate CSE summary scores were derived to serve as dependent variables, and selected WEI measures served as independent variables for part two of the study.

One categorical independent variable was determined based on responses to the nine items in the WEI
Job Factors section for the subject's current job. As the reader may recall, the total score for this section provides for classification of the job as entry, advanced, or administrative. Consequently, subgroups of the respondents to the WEI were formed indicating the level of responsibility in their current employment.

Continuous independent variables were derived for each of the remaining three sections of the WEI: (1) Knowledges, Skills and Abilities; (2) Position Responsibilities (job tasks and duties); and (3) Work Situations. The design of the WEI does not provide for categorization of subjects based on scores in these sections. Analyses to determine the relationship between each of these aspects of work experience and test performance were conducted. The objectives for part two of the study were as follows:

1. To determine if statistically significant differences in test performance exist between entry, advanced and administrative job level subgroups.

2. To determine the direction, magnitude and statistical significance of the relationship between the derived summary measure of test performance (dependent variable) and each of the following three independent variables:

   (a) WEI Knowledge, Skills and Abilities score
   (b) WEI Position Responsibilities score
   (c) WEI Work Situations score
The Clinical Simulation Examination

The Clinical Simulation Examination (CSE), as developed and administered by the National Board for Respiratory Care with technical assistance provided by Educational Testing Service (ETS), represents one portion of a two part examination for the advanced level respiratory care practitioner. Successful completion of the 100 item multiple-choice portion as well as the ten PMP clinical simulation portion are required in order for the Registered Respiratory Therapist (RRT) credential to be awarded. Both parts of the examination are offered on the same day at seventy ETS supervised test centers in the United States. The examination is administered twice per year, once in June and again in December. Following each administration of the examination, ETS scores the test booklets and provides separate item analyses and test statistics for the multiple-choice and written simulation portions of the exam.

Performance on the 100 item multiple-choice test was not included in the design of this study as subjects may not have taken that examination for one of many reasons. First, prior to January 1984, candidates for the CSE were not required to pass the 200 item entry level CRTT multiple-choice examination, but were rather required to pass a 200 item advanced level multiple-choice examination in order to become eligible for the
Therefore, some subjects may have taken the previously required longer advanced level multiple-choice test. Second, after January of 1984, candidates for the CSE were required to achieve the CRTT credential in order to become eligible for the CSE, and the advanced level multiple-choice examination was reduced to 100 items thereby eliminating content overlap with the 200 item CRTT examination; candidates may elect to take either one or both of the advanced level examinations on any test date. Therefore, some subjects may or may not have yet seated the revised 100 item advanced level multiple-choice examination. Lastly, subjects in the study sample are not necessarily eligible for any of the NBRC examinations and therefore, no previous test data may exist for some of these individuals.

Description

A description of the purpose and design of the Clinical Simulation Examination was provided in Chapter One. Additionally, a sample problem from the NBRC Candidate Handbook may be found in Appendix C. The problem pool, from which ten problems are selected according to strict decision rules to comprise any one form of the exam, consists of approximately 55 problems.

The aggregate of problems which comprise any one form of the CSE are never identical. To provide further
description, the following information is provided regarding the June 1984 CSE test form to be used in the study.

1. The aggregate of problems contains 26 IG sections and 92 DM sections; the minimum number a candidate will complete is a total of 98 sections (25 IG and 73 DM) which assumes that the most direct route is followed through each problem.

2. Each problem contains 1-4 IG sections (average = 2.6), and 8-10 DM sections (average = 9.2); the most direct route through each problem ranges from 8-11 total sections (average = 9.8).

3. The total number of options for the ten problems is 1021, with a range of 66-153 and an average of 102.1 per problem.

4. Although there are many more DM sections in the exam, the number of IG and DM options are about the same. In IG sections, the candidate selects as many as are desired from a list of 12-26 options; the total number of IG options is 495, with a range of 16-94 and a mean of 49.5 per problem. In DM sections, the candidate chooses only one option from a list of 4-10 (unless the response to a selected option instructs to choose another); the total number of DM options is 526, with a range of 42-63 and a mean of 52.6 per problem.

5. The distribution of problem difficulty levels is determined by calculating a P+ value for each problem (i.e., P+ equals the percentage of candidates passing the problem). For this exam, one problem has a P+ of less than .35, one greater than .65, and the remainder mid-range between .35 and .65.

6. Due to the difficulty with applying traditional methods of estimating reliability, reliability for the examination per se is not calculated. Rather, ETS provides a separate internal consistency estimate for the IG and DM responses for each route through the problem. These IG and DM "route" reliabilities range from .72 to 1.00 respectively.
Content Validity

The content validity of the CSE is demonstrated through standardized and rigorous test development procedures (Hixon, 1982). The content of the exam is based on data obtained from a national job analysis study completed by ETS for the NBRC (Nettles, 1982). A thirteen content area category-by-process (IG and DM) matrix delineates the scope for each CSE test form and additional specifications delineate the type of patient problems to be included on each form. In this way, the aggregate of problems which form each examination represent the scope of respiratory care practice in terms of patient type (neonatal, pediatric, adult), disease type (chronic vs. acute pulmonary, trauma-induced, cardiac, neuromuscular) and context of care (emergency, critical, or routine), as well as appropriate dispersion of test content delineated by the matrix. A copy of the June 1984 CSE test matrix may be found in Appendix B.

Problems are initially obtained from contracted CSE item writers who are RRT practitioners or educators that have attended an NBRC sponsored item writers workshop. The writer develops a PMP according to the criteria delineated in the contract; a writer declines the contract if he or she feels incapable of developing a problem according to the criteria specified. A completed problem is submitted to the NBRC Executive office for initial
review. If it meets the contract specifications, the writer is paid and the problem forwarded to ETS to be prepared (initially edited) for committee review.

The CSE committee consists of ten individuals, five physicians and five Registered Respiratory Therapists (RRT's), with representation from each of the five NBRC sponsoring organizations. During the first committee review, the design, content and routes through the problem are revised by a group consensus method. A second review takes place three to six months later and if major content changes are not indicated, the problem scoring is finalized and it is submitted to the active problem pool for inclusion in future tests. If content revision is required, the problem must be brought for a third review.

The CSE committee assembles two forms of the exam for administration each year. Every attempt is made to assemble forms that are as parallel as possible. Not only is consideration given to appropriate coverage of the content-by-process matrix and problem type criteria discussed above, but also problem lengths and difficulty levels are balanced.

**Weighting and Scoring**

CSE options may be weighted positive, negative or zero according to criteria delineated for each weight.
The upper limit of the scale is +3 which indicates an option of central importance for optimal patient care, a +2 indicates an option strongly facilitative and a +1, mildly facilitative of good patient care. A zero weight is used for options that are neither helpful or harmful, or for controversial options due to regional differences in care. The lower limit of the scale is -3 which indicates an option gravely damaging to patient care, a -2 indicates an option seriously detrimental and a -1, mildly detrimental to patient care in terms of cost, time, risk of morbidity or mortality.

Once the content of a problem has been finalized, every section is identified by the process to which it applies (IG or DM) and to which of the 13 content areas it applies (e.g., IG in clinical assessment category, or DM in ventilatory support category). Each option of the problem is then assigned a weight on the above scale by consensus of the CSE committee. Next, the minimum pass level (MPL) is determined for every section of the problem. This is also a judgmental process arrived at by group consensus wherein the committee decides what the score required to pass each section of the problem will be. The MPL for the problem is determined separately for IG and DM by summing the minimum pass values of all the IG and DM sections respectively (National Board for Respiratory Therapy, 1979).
When a set of problems are selected from the pool to form a test, the IG and DM minimum pass levels of the ten problems are summed to yield the IG and DM passing scores for the test. Following the test administration, the CSE committee reviews candidate question challenges and the item analysis results, and minor adjustments in the weighting of an option or a section MPL may be made if indicated. This process was completed for the June 1984 CSE with a resulting IG MPL of 131 (maximum possible points = 191) and DM MPL of 169 (maximum = 269), which converts to a 68.6% IG and a 62.8% DM proficiency score required to pass this form of the CSE.

CSE test booklets are scored by the use of a bar code wand reader which is connected to a programmed terminal. A unique bar code is printed in the test booklets directly adjacent to each response box which corresponds to a given option. The bar code identifies the problem number, the option number and weight, the content category of the option and the process to which it applies (IG or DM). Test booklets are manually scanned and the scorer brushes the wand reader over the bar code of each response which has been selected (i.e., exposed by the latent image developer pen). As a quality check, ten percent of the booklets are randomly selected and rescored.

Final score reports are generated from the computer
to which the bar code reader is connected. Scores are reported by problem, for the total test and by content category. A candidate's score in any of these areas is simply the sum of the weights of all options selected which correspond to the score category. For each problem, the candidate's observed IG score, the IG MPL score, and the IG maximum possible score is reported as is the observed DM score, the DM MPL and the DM maximum. The same is done for the ten problems as a whole indicating total test performance, and these total IG and DM test scores are also reported as a percentage of the maximum possible points to aid candidate understanding. Pass/fail status is determined solely on the basis of the total test scores, and one must achieve the passing score in both IG and DM to pass the test. To provide feedback to candidates on relative areas of strength and weakness, the cross-referencing of IG and DM sections to the thirteen content areas allows for content area scores to be generated. These are simply reported as the number of points achieved in a given process and content category across all ten problems. A sample CSE score report may be found in Appendix A.

**Criterion-related Validity**

A confidential criterion-related validity study of the CSE was voluntarily conducted by this researcher in
1984 in exchange for the use of data for dissertation purposes as previously described. The study utilized supervisor ratings of job performance as the criterion variable and IG and DM total test scores as predictor variables for the correlational analyses.

The study sample (also previously described) consisted of a geographically representative sample of respiratory care practitioners from twelve selected hospitals in the U.S. Each subject completed the June 1984 CSE test form in July or August under secure test conditions at their place of employment. As subjects in the sample may have already held the RRT credential, or at the opposite extreme, may have not met admission criteria for this exam, no credentialing decisions were based on test results. Rather, the test was administered solely for research purposes and the diversity in practitioner backgrounds was desirable in order to avoid restriction of range difficulties.

The clinical supervisor of each subject, having no knowledge of test results, provided thirty-one ratings of job performance on a standardized task-based instrument. Results of the study revealed statistically significant (p< .01) positive correlations between the composite job performance rating scores and IG and DM summary scores respectively, indicating that CSE scores are predictive of performance on the job (Bryant, 1985).
The Work Experience Inventory

The Work Experience Inventory (WEI) was developed and pilot tested by ETS for the NBRC in 1982. Although a one year clinical experience requirement is included in the admission criteria for the NBRC advanced practitioner (RRT) examination, no information is obtained about this experience and no systematic investigations have been conducted regarding the nature of such experience. However, national manpower studies have consistently demonstrated that RRTs hold higher level (i.e., more responsible) positions in the field; therefore, the purpose of the WEI development and pilot testing was to determine if a standardized self-report instrument could differentiate between RRTs and non-RRTs in terms of relevant work experience, and if the information was accurate (Nettles, 1984).

The instrument was developed by ETS with assistance from an NBRC advisory committee. The WEI provides data on many job characteristics including level of responsibility on the job, exposure to different work situations and job tasks, and to a variety of respiratory therapy knowledges, skills and abilities. The draft instrument consisted of the following sections:

1. Employment History - similar to a standard employer's application blank. The employer's name and address is requested here as is the subject's tenure and supervisor, and information about specific job responsibility areas.
2. Job Factors - subjects provide ratings for their current job, and may rate their two most recent previous jobs if applicable, on nine factors considered important for pay and classification of hospital positions (such as prior education, job complexity, supervision and consequences of errors) using a multiple-choice format; these items are weighted and the total points achieved in this section allows for classification of the job as entry, advanced, or administrative.

3. Work Situations - a list of 22 work situations is provided and the subject rates the degree to which each represents a significant part of past or present positions held, using a scale of zero (does not apply) to four (one of the most significant) inclusive; scores are reported as the raw rating on each of these items.

4. Position Responsibilities - 82 of the most important tasks identified by the advanced level job analysis study are listed and the subject indicates the degree to which each task was a significant part of past or present positions held using the same scale as above; sets of these items are rationally categorized into 25 job responsibility areas and the mean rating for the items defining each category is reported.

5. Knowledges, Skills and Abilities - the same rating procedure is used to indicate the significance in work experience of 62 items which are rationally categorized into 19 content areas; the mean rating for the items included in each category is reported.

6. Biographical and demographic information - such things as sex, race and type of hospital in which employed were gathered for subgroup analysis.

The draft WEI was pretested on a sample of ten RRTs and non-RRTs. The results and their comments were used to finalize the instrument. Since one problem with using self-report information is fabrication or inflation of responses, an "Employer's Verification Form" was also developed. This was an abbreviated version of the WEI that could be completed by employers in a short time and used to determine the accuracy of subject responses.
Subsequently, a stratified sample of practitioners were selected for pilot-testing of the instrument. Three hundred of the 1800 applicants to the June 1982 CSE were randomly sampled by geographic region to represent the non-RRT population; two hundred of the 5400 individuals currently holding the RRT credential were randomly sampled by geographic region to represent the RRT population. Each subject was mailed a package containing a cover letter from the NBRC Executive Director, a copy of the WEI and an Employer's Verification Form. Subjects were asked to complete the WEI and return it to ETS for scoring; subjects were also requested to ask their department director to complete the Verification Form and return it directly to ETS. Each subject was subsequently provided their WEI Profile, a type of score report using computer graphics to display results.

The WEI was returned by 138 subjects, a return rate of 27.6% overall; a 27% return rate for RRTs and 28% for non-RRTs. Ninety percent of both groups indicated that they felt the instrument provided an adequate means of describing their relevant work experience. Employer Verification Forms were returned for 126 of the 138 subjects; a return rate of 91%.

Subgroup analyses of mean differences in item ratings revealed few significant differences by sex or hospital in which employed, but many significant
differences between the RRT and non-RRT groups (p<.05). Additionally, intercorrelational analyses of item ratings by sections of the inventory demonstrated very few values greater than .50; thus the conclusion that there was little overlap in the WEI items and that each WEI section was contributing somewhat unique information.

When the NBRC advisory committee reviewed the results of the pilot testing analyses, they found them to be meaningful in terms of the instrument providing a reasonable method of quantifying work experience and differentiating between RRTs and non-RRTs (Nettles, 1984).

The results of the analyses using Employer Verification data were not so encouraging. Although there was some agreement between the employer and respondent for many of the verified items, the correlation between the total score for the abbreviated verification form and the respective subject's WEI score was .55 for the non-RRT group and .35 for the RRT group. It was felt that the information requested on the Employer's Verification Form may have adversely affected the accuracy analysis as only the Job Factors and Knowledges, Skills and Abilities items were included. That is, although a department director may be able to reasonably provide accurate information regarding job factors items, the knowledges, skills and abilities items clearly represent unobservable behaviors which would be difficult for the
employer to accurately rate (Nettles, 1984). This was admittedly an oversight in the verification form design.

Based on the recommendations of ETS consultants, the NBRC advisory committee made some minor changes to the WEI instrument (i.e., expanded the range of the rating scale to a maximum of seven). They also suggested that the NBRC, in cooperation with ETS, market the WEI in order that interested practitioners may complete the instrument and obtain their Profile report as a self-assessment exercise. The revised WEI is now available for a fee through the NBRC, but employer's verification data is not requested as this was deemed intimidating and therefore not appropriate to the use of the instrument for self-assessment purposes. A copy of the WEI which is now available for purchase, but was provided free of charge to the study sample for this research project, may be found in Appendix D. Also, a sample WEI Profile, the computer graphics WEI score report, is presented in Appendix E.

Data Analysis Procedures

This section provides an overview of the data analysis procedures established to meet the objectives of the study. Results of the statistical analyses are presented in Chapter Four and discussed in Chapter Five.
Part One: Description, Dimensionality and Reliability

Part One of the study involved an analysis of selected psychometric properties of the CSE using twenty score variables (10 IG and 10 DM problem scores) for two groups of subjects. One sample consisted of the 2821 NBRC candidates who completed the June 1984 CSE for credentialing purposes (termed the candidate sample), and the other consisted of the 312 subjects who completed the same test form later in 1984 solely for research purposes (termed the study sample). The candidate sample provided a sufficiently large number of subjects required for factor analysis procedures; the subject-to-variable ratio was approximately 140:1. The study sample, although considerably smaller, provided the only available data to compare results on the same test form across groups; the subject-to-variable ratio for this sample was approximately 15:1. The statistical analysis procedures planned for part one of the study are briefly described below.

Descriptive Statistics. The mean and standard deviation for each test score (dependent) variable was to be calculated for each group of subjects. In order to determine if the samples demonstrated significant differences in ability level, a one-way MANOVA was to be conducted for the twenty dependent variables; if the
results of this analysis indicated a significant difference between samples (Wilk's Lambda, p< .01), then a one-way ANOVA for each of the dependent variables was to be conducted to determine on which dependent variables the samples significantly differed (F test, p< .01).

**Factor Analyses.** The factor analysis procedures which are described below were conducted separately for each sample. Due to the larger number of subjects in the candidate sample and consequently less probability of obtaining chance results, each analysis was conducted for this sample initially, and repeated for the study sample for comparison purposes.

1. First, a 20 x 20 symmetrical correlation matrix for the twenty problem score variables was obtained. This matrix served as input to the LISREL confirmatory factor analysis program which employs a maximum likelihood procedure for analysis.

2. The data was analyzed for "goodness of fit" to a null model (no common factors), and to a hypothesized model (two common factors compatible with the dimensions of IG and DM). These results were compared by use of the rho statistic which provides a ratio of the improvement of fit between the null and the hypothesized model, compared to the correct or "perfect fit" model (Bentler and Bonett, 1980). The hypothesized two factor model, representing the dimensions of IG and DM respectively, was considered confirmed if the rho statistic approached 1.0; a value of .90 or greater was considered acceptable.

3. As an acceptable fit to the hypothesized two factor model was not obtained, the SAS principal axis exploratory factor analysis procedure was used to determine the factor structure of the
20 x 20 correlation matrix. Squared multiple correlations were specified as the initial communality estimates, a scree plot of eigenvalues obtained, and the discontinuity approach as well as the eigenvalue greater than one criteria used to determine the number of factors appropriate to retain. Based on the factor solution, the number and type of scores (as opposed to the IG and DM summary scores used by the NBRC) to describe performance on the CSE as a result of this analysis was then determined.

4. Lastly, factor analysis results for each sample were subjectively inspected to assess the stability of the factor structure across groups on the same test form.

Reliability. The Cronbach alpha measure of internal consistency was computed for the problem score variables comprising each factor resulting from the above analyses.

Part Two: Relationship of Work Experience

Part Two of the study involved an investigation of the relationship of work experience to test performance. Part two analyses utilized only CSE and WEI data for the WEI respondent group of the study sample. Based on the factor analysis results from part one of the study, a test score for each dimension of CSE performance was to be derived to serve as the dependent variable in each of the analyses; data from the four previously described sections of the WEI instrument provided the basis for deriving independent variables for these analyses. The planned analysis procedures for this part of the
study are briefly described below.

**Descriptive Statistics.** The mean and standard deviation for each independent and dependent variable were determined for descriptive purposes.

**Analysis of Variance.** Based on scores achieved in the WEI Job Factors section, subjects were categorized on their Profile as holding an entry, advanced or administrative level job, for their current as well as up to two previous jobs which they may have rated. Considering only the classification for their current job (as this was the position held at the time CSE test data was obtained), a one-way ANOVA was conducted to determine if statistically significant differences in test performance existed based on job level (F test, p< .01). As some subjects failed to complete the WEI Job Factors section, only 97 of the 112 respondents were categorized by job level; of these approximately equal numbers were categorized as holding entry and advanced level jobs, but only six were at the administrative level and consequently were not included in this analysis procedure. Therefore, the independent variable for the ANOVA consisted of two levels - entry and advanced.

**Correlational Analyses.** Initially, an intercorrelation matrix for the three continuous independent variables (the WEI Work Situations, Job Responsibilities,
and Knowledges, Skills and Abilities section scores), and the test score dependent variable(s) was generated. These zero order correlations were to be tested for statistical significance using a standard table of critical $r$ values for a two-tailed test (alpha level .01), in order to determine the direction, magnitude and statistical significance of the relationships between these quantified measures of work experience pertaining to different aspects of the job and CSE test performance, as well as to assess the degree of multicollinearity of the independent variables. Lastly, multiple regression analyses were conducted, utilizing the three WEI section scores as predictors and the derived CSE test score variable (as determined from part one) as the criterion; multiple correlation coefficients were also calculated as well as proportions of variance explained by combinations of the predictors for each analysis.
IV. RESULTS

This chapter presents the specific procedures and results obtained from the statistical analyses of data for the study; the discussion of results and conclusions of the study are presented in Chapter Five. This chapter consists of two major sections which correspond to the two sets of research questions and objectives of the study presented in Chapters One and Three respectively.

Part One: Description, Dimensionality and Reliability

Preliminary analyses for this part of the study included the calculation of descriptive statistics for the twenty CSE problem score variables for the candidate and study groups respectively, and analyses to determine if the samples significantly differed in ability level as evidenced by CSE performance. The analysis procedures of central interest in this part of the study involved the use of factor analyses to determine the dimensionality of the twenty CSE problem score variables for each sample, in such a way as to determine the stability of the twenty score factor structure across
two samples of examinees on the same test form. Lastly, the reliability of the variables comprising each factor for each group was calculated.

Results of the Descriptive Analyses

The mean and standard deviation (S.D.) for each of the ten information gathering (IG) and ten decision making (DM) problem score variables for the candidate and study samples respectively are presented in Table 4. The bottom of the table also includes the mean and standard deviation for the total IG and DM test scores; these total test scores are calculated simply by summing the ten problem scores in each category.

As can be observed from the table, the mean for the candidate sample is higher for each problem score variable and consequently for the total test scores as well, and the standard deviation is consistently lower for each score variable for the candidate sample. As previously discussed in Chapter Three, it was expected that the candidate sample was a more homogeneous group of subjects and would perform at an overall higher ability level because these subjects all met admission criteria for the examination (i.e., educational and clinical experience requirements) and were seating the test for credentialing purposes.
Table 4.  
Candidate and Study Sample Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Candidate (n=2821)</th>
<th>Study (n=312)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>IG1</td>
<td>39.96</td>
<td>5.39</td>
</tr>
<tr>
<td>DM1</td>
<td>6.49</td>
<td>1.93</td>
</tr>
<tr>
<td>IG2</td>
<td>31.91</td>
<td>4.58</td>
</tr>
<tr>
<td>DM2</td>
<td>10.06</td>
<td>3.92</td>
</tr>
<tr>
<td>IG3</td>
<td>15.83</td>
<td>2.16</td>
</tr>
<tr>
<td>DM3</td>
<td>19.77</td>
<td>4.51</td>
</tr>
<tr>
<td>IG4</td>
<td>5.11</td>
<td>1.23</td>
</tr>
<tr>
<td>DM4</td>
<td>11.68</td>
<td>3.54</td>
</tr>
<tr>
<td>IG5</td>
<td>11.85</td>
<td>2.02</td>
</tr>
<tr>
<td>DM5</td>
<td>12.59</td>
<td>3.00</td>
</tr>
<tr>
<td>IG6</td>
<td>11.77</td>
<td>2.07</td>
</tr>
<tr>
<td>DM6</td>
<td>9.51</td>
<td>1.46</td>
</tr>
<tr>
<td>IG7</td>
<td>10.05</td>
<td>2.38</td>
</tr>
<tr>
<td>DM7</td>
<td>24.01</td>
<td>4.53</td>
</tr>
<tr>
<td>IG8</td>
<td>31.22</td>
<td>3.00</td>
</tr>
<tr>
<td>DM8</td>
<td>12.14</td>
<td>2.80</td>
</tr>
<tr>
<td>IG9</td>
<td>26.60</td>
<td>3.32</td>
</tr>
<tr>
<td>DM9</td>
<td>13.41</td>
<td>2.83</td>
</tr>
<tr>
<td>IG10</td>
<td>29.58</td>
<td>5.98</td>
</tr>
<tr>
<td>DM10</td>
<td>9.19</td>
<td>2.72</td>
</tr>
<tr>
<td>TOTAL IG</td>
<td>213.86</td>
<td>19.59</td>
</tr>
<tr>
<td>TOTAL DM</td>
<td>128.12</td>
<td>15.93</td>
</tr>
</tbody>
</table>
On the other hand, approximately one-third of the study sample subjects did not meet the educational and/or experience admission requirements for this examination, and therefore it was anticipated that greater variation in test scores would result for this sample. Also, the study sample completed the examination solely for research purposes and presumably would not prepare for and seat the test with the same motivation as would candidates desiring to achieve a credential.

In order to determine if statistically significant differences in performance existed across all twenty problem scores variables between groups, the null hypothesis that both group centroids (mean vectors) are equal was tested utilizing a multivariate analysis of variance (MANOVA). The alternative hypothesis for this analysis was that both group centroids are not equal.

A one-way MANOVA was completed using the SAS computer package, specifying the independent variable as group with two levels (candidate and study) and the twenty problem scores as dependent variables. The Wilks Lambda statistic, which compares the "within-groups" sums of squares cross products matrix to the "between-groups" sums of squares cross products matrix, was used as the test of significance for the multivariate hypotheses. A statistically significant difference resulted at the \( p < .0001 \) level indicating an overall significant
difference in performance between the two groups. Results of this analysis are presented below in Table 5.

Table 5.

| MANOVA by Group for the Twenty Problem Score Variables |
|-----------------|-----------------|-----------------|-----------------|
| Source          | df              | $F$             | Wilks' Lambda   | $p$ value       |
| Group           | 20/3112         | 12.27           | 0.9270          | $p < .0001$     |

In order to determine on which of the dependent variables the groups significantly differed, a series of one-way analysis of variance (ANOVA) procedures were completed; for each ANOVA, the independent variable was again specified as group with two levels but the dependent variable in each analysis was one of the twenty problem scores. The results of these analyses are presented in Table 6.

The $F$ statistic was significant at the $p < .01$ level for each of the score variables with the exception of the DM score on problem six (DM6). The error term for each ANOVA model contained 3131 degrees of freedom due to the large sample sizes of the groups (i.e., candidate $n=2821$, study $n=312$). Therefore, it is not surprising that statistically significant differences resulted for nearly all the problem score variables.
Table 6.
Results of Each One-Way Anova by Group for the Twenty Problem Score Variables

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>df</th>
<th>Mean Square</th>
<th>F value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG1</td>
<td>1</td>
<td>3623.4075</td>
<td>112.12  *</td>
</tr>
<tr>
<td>DM1</td>
<td>1</td>
<td>65.7703</td>
<td>16.78   *</td>
</tr>
<tr>
<td>IG2</td>
<td>1</td>
<td>2870.6979</td>
<td>120.04  *</td>
</tr>
<tr>
<td>DM2</td>
<td>1</td>
<td>128.7828</td>
<td>8.11    *</td>
</tr>
<tr>
<td>IG3</td>
<td>1</td>
<td>420.4583</td>
<td>79.25   *</td>
</tr>
<tr>
<td>DM3</td>
<td>1</td>
<td>1638.2803</td>
<td>73.54   *</td>
</tr>
<tr>
<td>IG4</td>
<td>1</td>
<td>42.5304</td>
<td>26.12   *</td>
</tr>
<tr>
<td>DM4</td>
<td>1</td>
<td>228.0939</td>
<td>17.46   *</td>
</tr>
<tr>
<td>IG5</td>
<td>1</td>
<td>127.5329</td>
<td>30.02   *</td>
</tr>
<tr>
<td>DM5</td>
<td>1</td>
<td>230.8390</td>
<td>24.36   *</td>
</tr>
<tr>
<td>IG6</td>
<td>1</td>
<td>323.0204</td>
<td>68.23   *</td>
</tr>
<tr>
<td>DM6</td>
<td>1</td>
<td>1.0107</td>
<td>0.45</td>
</tr>
<tr>
<td>IG7</td>
<td>1</td>
<td>156.4335</td>
<td>26.88   *</td>
</tr>
<tr>
<td>DM7</td>
<td>1</td>
<td>833.5603</td>
<td>37.58   *</td>
</tr>
<tr>
<td>IG8</td>
<td>1</td>
<td>549.1983</td>
<td>56.11   *</td>
</tr>
<tr>
<td>DM8</td>
<td>1</td>
<td>100.3474</td>
<td>12.24   *</td>
</tr>
<tr>
<td>IG9</td>
<td>1</td>
<td>364.6650</td>
<td>30.38   *</td>
</tr>
<tr>
<td>DM9</td>
<td>1</td>
<td>313.3410</td>
<td>36.38   *</td>
</tr>
<tr>
<td>IG10</td>
<td>1</td>
<td>1721.7025</td>
<td>45.73   *</td>
</tr>
<tr>
<td>DM10</td>
<td>1</td>
<td>316.0902</td>
<td>39.87   *</td>
</tr>
</tbody>
</table>

* p< .01
Nevertheless, the null hypothesis for the MANOVA must be rejected in favor of the alternative hypothesis; that is, the group centroids are not equal indicating that the two samples are probably not from the same population. Additionally the follow-up one-way ANOVAs indicate that statistically significant differences in performance were evident for all of the dependent (problem score) variables with the single exception of the DM6 score. It should be noted that the direction of the difference in each case is that the candidate sample performed at a higher level than the study sample.

A second MANOVA was conducted to determine if statistically significant differences existed in overall test performance as evidenced by the total IG and DM test scores. As would be expected, the Wilks' Lambda statistic for this analysis using total IG and DM test scores as dependent variables was 0.9397 (df=2/3130) and significant at the p< .0001 level. Follow-up one-way ANOVAs for the total test IG and DM score variables respectively resulted in F values of 163.09 and 156.39, both of which are statistically significant at the p< .01 level. This is also as expected because the total test score variables are merely simple sums of the problem score variables in each category (IG and DM).

Lastly, in terms of investigating if the samples differed in ability level based on CSE test performance,
examination pass rates for the two samples were determined. It was found that 50% of the subjects in the candidate sample passed the test (i.e., achieved at or above the minimum pass level set for both IG and DM on this form of the CSE) which is similar to other groups of candidates who have seated this examination in recent years (i.e., passing percentages are commonly in the 50-60% range); however, subjects in the study sample passed at the considerably lower rate of 33%. Thus, these results appear to indicate, as was expected, that the candidate sample demonstrated a higher ability level on the test than the study sample.

**Factor Analysis Procedures and Results**

Factor analyses were completed separately for the candidate and study samples. Each factor analysis was completed for the candidate sample initially, then repeated with the study sample data for comparison purposes. It was felt that this was the best approach as the candidate sample consisted of a larger sample (i.e., n=2821 versus n=312), and therefore provided a larger subject-to-variable ratio (i.e., approximately 140:1 versus 15:1) for the analyses; consequently, there was a lesser probability of obtaining chance results for the candidate sample.

**Confirmatory Factor Analyses.** The first factor
analysis procedure involved determining the goodness-of-fit of the ten IG and ten DM problem scores to the hypothesized two factor model. As the reader will recall, previous factor analysis studies of measures derived from physician PMP tests have demonstrated a two factor structure consisting of an information or data gathering dimension and a decision making dimension (Donnelly, Gallagher, Hess and Hogan, 1974; Juul, Noe and Nerenberg, 1979; Skakun, 1978). Additionally, the NBRC examination was specifically developed according to this two dimensional model and therefore, confirmatory factor analysis was deemed the procedure of choice.

Confirmatory factor analysis is a method for examining the measurement aspects of structural models and offers several advantages over traditional exploratory factor analysis procedures; the primary advantage of confirmatory factor analysis is that it provides for hypothesis testing by enabling the researcher to state a theory or hypothesis in the form of a statistical model which specifies the links between the unobserved or latent variables (underlying factors) and the observed measures (Long, 1983; Joreskog, 1979). The specified model is then tested with actual data to determine the goodness-of-fit of the data to the model.

The LISREL confirmatory factor analysis program, which employs a maximum likelihood procedure for analysis
was used for this portion of the study. The basic measurement model incorporated in this program is the population version of the common factor model. That is, the maximum likelihood procedure is based on the assumption that the sample from which the data was obtained represents some defined population.

Specifying the statistical model for confirmatory factor analysis involves designating parameter values as fixed (i.e., set at 0 or 1) or free (i.e., allowed to vary) for each of three "target" matrices, such that the model corresponds to the factor structure hypothesized to represent the data. It is by designating certain parameter values that one defines the theory or hypothesized model; if all parameter values were left free, it would be equivalent to exploratory factor analysis. Parameter values for the following three matrices are designated as free or fixed in order to specify the statistical model for analysis: (1) the lambda matrix which represents the hypothesized population factor loadings, (2) the phi matrix which represents the hypothesized correlation among the factors in the population, and (3) the theta matrix which represents the unique variances of the variables (Joreskog and Sorbom, 1983).

The maximum likelihood procedure employed in the LISREL program utilizes iteration to maximize the
likelihood function for which the input data best fits the specified model; rotation is therefore not applicable to this factor analysis method (i.e., simple structure is "built into" the model). The LISREL program produces a chi-square statistic which indicates the goodness-of-fit of the input data to the hypothesized model. Unlike traditional uses of the chi-square test statistic, in this application the preferred outcome is a nonsignificant chi-square value. That is, the hypothesis being tested in this application is that there are no significant differences between the specified model and the fit of the data to it. Consequently, a nonsignificant chi-square would indicate that the factor structure of the input data does not significantly differ from the hypothesized model, which is the desired outcome in confirmatory factor analysis.

However, the chi-square statistic is often inconclusive because the hypothesis being tested is very stringent (i.e., the model specified holds exactly in the population), and with the large sample sizes commonly utilized in factor analysis, the chi-square statistic will virtually always be significant leading to the conclusion that the data does not adequately reflect the hypothesized model. Therefore, Bentler and Bonett (1980) have proposed an alternative goodness-of-fit measure which is an adaptation of the Tucker-Lewis
coefficient and called the "rho" statistic. Calculation of this statistic requires specifying a model for an initial LISREL analysis which indicates that there are no common factors (i.e., the variables are uncorrelated in the population); this is referred to as the null model and represents the worst possible model for the data. The chi-square value for this null model analysis is then used as a baseline for comparison in calculation of the rho statistic. That is, a second LISREL analysis is obtained for the same data set with the model specified according to the hypothesized theory, and the chi-square value is thus obtained for what is referred to as the hypothesized model. These two chi-square values, and their respective degrees of freedom (determined based on the number of variables in the data set and the number of parameters in the specified model which are left free to vary), are used to calculate the rho goodness-of-fit statistic.

The rho statistic provides a ratio of the improvement of fit going from the null to the hypothesized model, compared to the correct or "perfect fit" model. The rho statistic is calculated as follows:

\[ \rho = \frac{Q_0 - Q_a}{Q_0 - 1} \]

where \( Q_0 \) = chi-square/df for the null model analysis, and \( Q_a \) = chi-square/df for the hypothesized model.

The maximum value of \( \rho \) is one (1=perfect fit); a
value of .90 or greater is generally considered acceptable as an adequate fit of the data to the hypothesized model. Due to the recognized inconclusiveness of the chi-square statistic in confirmatory factor analysis applications, the rho statistic was used as the definitive goodness-of-fit measure for this study.

A 20 x 20 symmetrical correlation matrix of the twenty problem score variables was generated for the candidate and study samples respectively. These matrices are reproduced in Appendix F. All intercorrelations for the candidate sample are positive and statistically significant at the p< .01 level; similarly, all intercorrelations for the study sample are positive and statistically significant at the p< .01 level, with the exception of a few. Also, many of the study sample intercorrelations are noted to be slightly higher than those for the candidate sample, as may be expected due to the greater score variability of the study sample.

Two LISREL analyses were performed on each data set; one specifying the null model (no common factors) and one specifying the hypothesized model (two common factors in which the ten IG score variables defined Factor 1 and the ten DM score variables Factor 2). The chi-square value and the level of significance for the hypothesized model analysis for each data set are presented in Table 7. Both indicated significant
differences between the hypothesized model and the respective input data set, but for reasons previously described, were considered inconclusive evidence on which to determine acceptable fit of the data to the model.

Table 7. LISREL Chi-Square Goodness-of-Fit Values for the Candidate and Study Data Sets

<table>
<thead>
<tr>
<th>Data Set (sample)</th>
<th>Chi-square</th>
<th>df</th>
<th>p level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate</td>
<td>1953.76</td>
<td>169</td>
<td>0.000</td>
</tr>
<tr>
<td>Study</td>
<td>518.01</td>
<td>169</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Using the chi-square information from the LISREL null and hypothesized model analyses for each data set, the rho goodness-of-fit statistic was calculated to conclusively determine if either of the data sets provided an acceptable fit to the hypothesized two factor model. The rho statistic results are presented in Table 8. The rho value for each data set was well below the minimally acceptable level of .90, indicating that neither data set provided an adequate representation of the hypothesized two factor model. Therefore, it was concluded that the hypothesized two factor structure could not be confirmed for either sample.
Table 8. Rho Statistic Goodness-of-Fit Values for the Candidate and Study Data Sets

<table>
<thead>
<tr>
<th>Data Set (sample)</th>
<th>Null Model Chi-square/df</th>
<th>Hypothesized Model Chi-square/df</th>
<th>Rho Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate</td>
<td>9269.82/190</td>
<td>1953.76/169</td>
<td>.7790</td>
</tr>
<tr>
<td>Sample</td>
<td>2305.72/190</td>
<td>518.01/169</td>
<td>.8145</td>
</tr>
</tbody>
</table>

Consequently, exploratory factor analyses were conducted to determine the dimensionality of the twenty score variables for each sample.

**Exploratory Factor Analyses.** The candidate and study sample correlation matrices were individually submitted to the SAS exploratory factor analysis program. The principal axis factor analysis procedure was used and squared multiple correlations were specified as initial communality estimates. A scree plot of eigenvalues was obtained for each of the data sets to determine the number of factors to be retained for analysis. Figures 3 and 4 are reproductions of these scree plots. Inspection of the eigenvalue plots (discontinuity approach) clearly indicates that one factor should be retained for the analysis of both the candidate and study data sets.
Figure 3. SAS Scree Plot of Eigenvalues for the Candidate Data Set
Figure 4. SAS Scree Plot of Eigenvalues for the Study Data Set
Additionally, use of the eigenvalue greater than one criteria also indicated that a one factor solution was appropriate for both data sets. The cut-off of one was not arbitrary in either case as the first eigenvalue is much greater than one, and the second eigenvalue considerably less than one for both data sets. Also, the first eigenvalue in each case accounts for more than eighty percent of the variance, with the second eigenvalue accounting for considerably less variance (i.e., between ten and thirteen percent). Therefore, it was determined that the appropriate number of factors to be retained for subsequent analysis in each case was one. Table 9 presents the eigenvalues and percent of variance explained for each data set.

Subsequent analysis with the principal axis factor procedure, specifying one as the number of factors to be retained and squared multiple correlations as the prior communality estimates, resulted in factor structures with similar loadings for each data set. The factor loadings for each data set are presented in a highest to lowest order in Table 10. Although the loading order of the variables is not the same for both data sets, the loading orders are very similar with the majority of IG variables loading highest on the factor. Also, it should be noted that all variables are positively correlated with the factor in each solution.
Table 9. Eigenvalues and Percent of Variance Explained for the Candidate and Study Data Sets

<table>
<thead>
<tr>
<th>Factor</th>
<th>Candidate Eigenv</th>
<th>Percent</th>
<th>Candidate Eigenv</th>
<th>Percent</th>
<th>Study Eigenv</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.7255</td>
<td>86.17</td>
<td>6.5192</td>
<td>80.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.5442</td>
<td>12.59</td>
<td>0.8415</td>
<td>10.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.4556</td>
<td>10.54</td>
<td>0.5218</td>
<td>6.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.3153</td>
<td>7.29</td>
<td>0.3867</td>
<td>4.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.2143</td>
<td>4.96</td>
<td>0.3404</td>
<td>4.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.2009</td>
<td>4.65</td>
<td>0.3082</td>
<td>3.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.1506</td>
<td>3.48</td>
<td>0.2772</td>
<td>3.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.1218</td>
<td>2.82</td>
<td>0.1347</td>
<td>1.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.0993</td>
<td>2.30</td>
<td>0.1011</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.0666</td>
<td>1.54</td>
<td>0.0826</td>
<td>1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-0.0163</td>
<td>-0.38</td>
<td>0.0135</td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>-0.0597</td>
<td>-1.38</td>
<td>-0.0222</td>
<td>-0.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>-0.1209</td>
<td>-2.80</td>
<td>-0.0632</td>
<td>-0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>-0.1493</td>
<td>-3.45</td>
<td>-0.1145</td>
<td>-1.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-0.1624</td>
<td>-3.76</td>
<td>-0.1311</td>
<td>-1.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>-0.1830</td>
<td>-4.23</td>
<td>-0.1753</td>
<td>-2.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>-0.1884</td>
<td>-4.36</td>
<td>-0.1977</td>
<td>-2.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>-0.2096</td>
<td>-4.85</td>
<td>-0.2030</td>
<td>-2.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>-0.2331</td>
<td>-5.39</td>
<td>-0.2447</td>
<td>-3.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-0.2478</td>
<td>-5.73</td>
<td>-0.2594</td>
<td>-3.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10. Factor Loadings of the one Factor Solution for the Candidate and Study Data Sets

<table>
<thead>
<tr>
<th>Candidate Variable</th>
<th>Loading</th>
<th>Study Variable</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>IG8</td>
<td>.657</td>
<td>IG2</td>
<td>.744</td>
</tr>
<tr>
<td>IG1</td>
<td>.581</td>
<td>IG8</td>
<td>.726</td>
</tr>
<tr>
<td>IG2</td>
<td>.577</td>
<td>IG6</td>
<td>.692</td>
</tr>
<tr>
<td>IG3</td>
<td>.573</td>
<td>IG9</td>
<td>.687</td>
</tr>
<tr>
<td>IG9</td>
<td>.554</td>
<td>IG3</td>
<td>.666</td>
</tr>
<tr>
<td>IG6</td>
<td>.508</td>
<td>IG1</td>
<td>.659</td>
</tr>
<tr>
<td>IG10</td>
<td>.468</td>
<td>IG10</td>
<td>.613</td>
</tr>
<tr>
<td>IG5</td>
<td>.457</td>
<td>DM4</td>
<td>.592</td>
</tr>
<tr>
<td>IG7</td>
<td>.419</td>
<td>DM3</td>
<td>.565</td>
</tr>
<tr>
<td>DM10</td>
<td>.400</td>
<td>DM10</td>
<td>.554</td>
</tr>
<tr>
<td>DM3</td>
<td>.392</td>
<td>DM7</td>
<td>.538</td>
</tr>
<tr>
<td>DM1</td>
<td>.369</td>
<td>IG5</td>
<td>.529</td>
</tr>
<tr>
<td>DM4</td>
<td>.349</td>
<td>DM5</td>
<td>.523</td>
</tr>
<tr>
<td>IG4</td>
<td>.338</td>
<td>DM9</td>
<td>.509</td>
</tr>
<tr>
<td>DM2</td>
<td>.325</td>
<td>DM8</td>
<td>.501</td>
</tr>
<tr>
<td>DM7</td>
<td>.315</td>
<td>IG4</td>
<td>.487</td>
</tr>
<tr>
<td>DM5</td>
<td>.273</td>
<td>IG7</td>
<td>.486</td>
</tr>
<tr>
<td>DM8</td>
<td>.264</td>
<td>DM6</td>
<td>.419</td>
</tr>
<tr>
<td>DM6</td>
<td>.223</td>
<td>DM2</td>
<td>.374</td>
</tr>
<tr>
<td>DM9</td>
<td>.208</td>
<td>DM1</td>
<td>.343</td>
</tr>
</tbody>
</table>
However, each of the loadings for the study sample are at least slightly higher than the same variable loading for the candidate sample with one exception; the DM1 loading is slightly less for the study sample (.343) than for the candidate sample (.369). The one factor solution accounts for at least ten percent of the variance for each variable in the study data set (i.e., all factor loadings are at least .343 so when squared, equal greater than .10 or ten percent); however, the one factor solution fails to account for at least ten percent of the variance for the four lowest variable loadings in the candidate data set. Nevertheless, each factor loading in both solutions represents a statistically significant correlation between the variable and the factor at the p< .01 level for a two-tailed test.

In summary, both data sets failed to provide an acceptable goodness-of-fit to the hypothesized two factor model and therefore, results of the LISREL confirmatory factor analysis for both samples must be interpreted as failing to confirm the hypothesis that the IG and DM score variables represent two respective dimensions underlying test performance on the CSE. Additionally, the exploratory principal axis factor analyses resulted in consistent one factor solutions for the samples with similar amounts of variance explained by each solution, and factor loadings of somewhat similar magnitude and
order with the majority of the IG variables loading highest on the factor. These one factor solutions are noted to be at variance with the results of Donnelly et al. (1974), Juul et. al. (1979), and Skakun (1978) whose studies of measures derived from physician PMP tests demonstrated two factor solutions.

**Cronbach Alpha Reliability Estimates**

The Cronbach Alpha measure of internal consistency was calculated for the variables comprising the one factor resulting from each exploratory factor analysis solution to provide an estimate of reliability. The resulting reliability coefficients for the twenty problem score variables were 0.78 and 0.88 for the candidate and study samples respectively. The greater variability in the study sample scores contributed to the higher coefficient for this group, but both values were considered acceptable reliability estimates for such a PMP test in which the focus of the scoring is not to identify individual differences in performance levels, but rather to identify group differences (i.e., passing versus failing performance).
Part Two: Relationship of Work Experience

The analysis procedures of central interest in this part of the study involved determining the relationship between measures of work experience derived from the WEI instrument and the CSE factor scores. This portion of the study used data from that subset of the study sample which returned a completed WEI instrument; these subjects comprised the respondent group (n=112).

Based on the one factor principal axis solution resulting from part one of the study and the twenty raw CSE problem score variables, a SAS standardized factor score was determined for each subject in the study sample. Upon return of the WEI instruments, each respondent's derived CSE factor score was obtained from the print-out and matched with their WEI scores. The CSE factor score provided a summary measure of test performance which could be utilized as the dependent variable for part two analyses.

Preliminarily, descriptive statistics for the WEI score variables and the derived factor scores were obtained for this respondent group. In each subsequent analysis, one or more of the WEI measures served as independent variables and the derived factor score as the dependent variable. A one-way ANOVA was conducted to determine if statistically significant differences in test performance existed based on the WEI Job Factors
classification of subjects as currently holding an entry or an advanced level job. Also, correlational analysis methods were employed to determine the relationship of the three quantified WEI measures (i.e., Work Situations, Job Responsibilities, and Knowledges, Skills and Abilities section scores) to CSE test performance.

Results of the Descriptive Analyses

The mean and standard deviation (S.D.) for each of the three quantified WEI scores and the derived CSE factor score for the 112 respondents are presented in Table 11 below.

Table 11.

Descriptive Statistics for the WEI Respondent Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSE Standardized Factor Score</td>
<td>0.32</td>
<td>0.76</td>
</tr>
<tr>
<td>WEI Work Situations</td>
<td>59.14</td>
<td>20.83</td>
</tr>
<tr>
<td>WEI Job Responsibilities</td>
<td>107.63</td>
<td>25.43</td>
</tr>
<tr>
<td>WEI Knowledges, Skills, and Abilities</td>
<td>87.54</td>
<td>22.75</td>
</tr>
</tbody>
</table>
As can be observed from the table, the average standardized CSE factor score for the respondent group is slightly above the mean of the total study sample of which it is a subset (i.e., total study sample standardized score mean = 0), and the standard deviation is slightly lower than that of the total study sample (i.e., standardized standard deviation = 1). The mean for Work Situations is noted to be lower with respect to the maximum score for this section, as compared to the other two WEI section score means compared to their maximums. That is, the maximum score for Work Situations = 154, Job Responsibilities maximum = 175, and Knowledges, Skills and Abilities maximum = 133. This appears to indicate that a ceiling effect was not prevalent in the self-ratings, and the rather large standard deviations for each WEI score variable indicate no restriction of range problem. Scatterplots of the three WEI score variables also displayed reasonable linearity.

Results of the One-Way ANOVA by Job Level

As the reader may recall, the WEI Job Factors section provided the respondent an opportunity to rate their current job, as well as two previous jobs, on nine factors considered important for pay and classification of hospital positions in a multiple-choice format. Responses to these nine weighted items result in a total
score which forms the basis upon which each job rated is categorized as entry, advanced or administrative. For purposes of this study, only the job level category for each respondent's current job was considered as this was the position held at the time of completing the CSE.

As some respondents did not complete the Job Factors section, only ninety-seven subjects could be categorized by job level. Of these, 43 were currently classified as entry, 48 as advanced and 6 as administrative job level. Therefore, it was decided that the small administrative group should not be considered, and the one-way ANOVA by job level was conducted using only the remaining 91 subjects. The dependent variable for the analysis was the CSE standardized factor score and the independent variable was specified as job level which contained two groups, entry and advanced. The results of the ANOVA are presented below in Table 12.

Table 12.
Results of the One-Way ANOVA for CSE Score by Entry and Advanced Job Level

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
<th>F value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1</td>
<td>5.1310</td>
<td>9.79 *</td>
</tr>
<tr>
<td>Within Groups</td>
<td>89</td>
<td>0.5241</td>
<td></td>
</tr>
</tbody>
</table>

* p< .01
The resulting F value was statistically significant at the p< .01 level, indicating that significant differences in CSE test performance did exist based on entry versus advanced level job classification. Inspection of the group means revealed that the advanced level group performed significantly better than the entry level group; the advanced level standardized test score mean = 0.5874 and the entry level mean = 0.1117.

**Correlational Analysis Procedures and Results**

A summary score for the WEI Work Situations, Job Responsibilities and Knowledges, Skills and Abilities section was respectively derived by summing the ratings within each category from the WEI Profile score report. Initially, an intercorrelation matrix was obtained which contained these three WEI score variables and the CSE factor score variable for inspection purposes. These intercorrelation are presented in Table 13.

As can be seen from the first column of intercorrelations on the table, each WEI variable is positively correlated with the CSE factor score and all are statistically significant at the p< .01 level for a two-tailed test. However, all of these correlations are quite low accounting for approximately 6-13 percent of the CSE score variance.
Table 13.
Intercorrelations for the Three WEI Score Variables and the CSE Factor Score Variable

<table>
<thead>
<tr>
<th></th>
<th>CSE Score</th>
<th>WEI Work Sit.</th>
<th>WEI Job Respons.</th>
<th>WEI KSAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSE Factor Score</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEI Work Situations</td>
<td>0.37 *</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEI Job Responsibilities</td>
<td>0.26 *</td>
<td>0.67 *</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>WEI Knowledges, Skills, Abilities</td>
<td>0.25 *</td>
<td>0.64 *</td>
<td>0.73*</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* p < .01

On the other hand, the WEI score variables are noted to be very highly correlated with each other, accounting for 41 to 53 percent of the variance in each other. Clearly, the relationships are much stronger between and among the measures of work experience, than the relationship between any one of the work experience measures with the CSE factor score.

Nevertheless, a series of multiple regression analysis procedures were conducted specifying the CSE factor score as the dependent variable, and the three WEI section scores as independent variables. Due to the high degree of multicollinearity of the independent variables
the results of the stepwise selection, forward entry and backward elimination methods all resulted in only one variable entering/remaining in the multiple regression equation. As would be expected, the single predictor variable was Work Situations, as it had the highest zero-order correlation with the dependent variable. Results of the multiple regression analyses are presented below in Table 14.

Table 14.
Results of the Multiple Regression Analysis of WEI Section Scores on the CSE Factor Score

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean Square</th>
<th>F value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>8.5270</td>
<td>17.09 *</td>
</tr>
<tr>
<td>Residual</td>
<td>110</td>
<td>54.9000</td>
<td></td>
</tr>
</tbody>
</table>

* p< .01

Therefore, it was concluded that the WEI Work Situations score is most predictive of the CSE factor score, and this relationship is statistically significant at the p< .01 level. However, because the Work Situations score is the only variable in the equation, the multiple regression coefficient is equivalent to the zero-order correlation between these two variables (.366), and this accounts for only 13.4 percent of the variance in the CSE factor score.
V. DISCUSSION AND CONCLUSIONS

The first portion of this chapter includes a discussion of the part one data analysis results and conclusions, and describes further analysis procedures which were undertaken to attempt to clarify the substantive meaning of those results. The next portion of this chapter provides a discussion of the results and conclusions for part two of the study. The chapter concludes with a summary of the study and conclusions, as well as recommendations for future research.

Part One Discussion and Conclusions

The first objective for part one of the study was to determine if statistically significant differences in test performance existed between the candidate and study samples. Two samples of subjects who had completed the same form of the CSE were desirable in order to allow for comparison of factor analysis results for generalizability across groups, and it was not possible to obtain comparable measures of performance for two actual NBRC candidate groups as each form of the CSE
is somewhat unique; that is, two CSE forms never contain the same set of problems. Therefore, as test data was available for a sample who had completed the same CSE form as an NBRC candidate group eight weeks earlier, this data set was used for comparison purposes. It was recognized however, that there were substantive differences between these two groups, both in composition and test taking conditions.

**Candidate and Study Sample Differences**

The candidate sample was self-selected in that subjects had voluntarily applied for the NBRC examination with the goal of achieving the Registered Respiratory Therapist (RRT) credential, and all had met admission requirements for the examination which include successful completion of the NBRC Certified Respiratory Therapy Technician (CRTT) multiple-choice examination as well as documentation of a required amount of education and work experience. Therefore, this sample was considered representative of only that subset of the respiratory care practitioner population who meet these requirements and elect to seat the CSE.

On the other hand, the study sample had been previously selected without regard for test admission requirements, to complete the CSE solely for research purposes such that the criterion-related validity of
the test could be investigated. This sample was considered representative of the more global population of practitioners who comprise the respiratory therapy labor force in the U.S. The study sample included a large proportion of subjects who had previously achieved the RRT credential by passing the CSE or the clinical oral which preceded it, as well as a lesser proportion of subjects who were currently eligible for the CSE, and subjects who did not qualify in that they failed to meet CSE admission requirements. Additionally, subjects in the study sample completed the test at their place of employment rather than at established test centers, and as pass/fail status and credentialing decisions were not to be determined based on results, presumably these test conditions would not produce the same degree of motivation to perform optimally as would be the case for the candidate sample who seated the test for credentialing purposes.

Results of the descriptive analysis for the twenty CSE problem score variables, as well as the total IG and DM test scores, revealed a higher mean score and a lower standard deviation for the candidate sample as compared to the study sample, for each of the score variables. This was not surprising for the reasons described above, and inspection of each sample's score frequency distributions revealed a wider range of scores for the study
sample than that of the candidate sample, presumably with the outlying scores representing the extremes of subjects in the study sample who had previously achieved the RRT credential versus those not eligible for the CSE in terms of admission requirements. This resulted in a much greater variation for the study sample on each score variable, and combined with the differences in test conditions, resulted in lower score means as well.

Consequently, the MANOVA of the twenty problem score variables by test group demonstrated a statistically significant difference in test performance levels (Wilk's Lambda, p< .0001), indicating that the samples were not from the same population; a second MANOVA by group for the total test IG and DM scores resulted in the same finding. Follow-up one-way ANOVAs for each score variable revealed statistically significant differences for all of the score variables with the single exception of the decision making score on problem six (DM6); in each case of statistically significant results, the direction of the difference was that the candidate sample demonstrated a higher ability level.

It was felt by the researcher that these findings could be a function of the differences in composition of the two samples, chiefly due to the subgroup of subjects in the study sample who were not eligible for the CSE. In order to explore this possibility, total test score
descriptive statistics were obtained for those study subjects who already held the RRT credential or were eligible for the CSE (n=206), and those who did not meet test admission requirements (n=106). The RRT and eligible study subgroup had a total IG score mean of 215.58 and standard deviation of 17.81, which was not significantly different from the candidate sample total IG mean of 213.86 and standard deviation of 19.59. Similarly, this study sample subgroup did not demonstrate significant differences as compared to the candidate sample for the total DM score values; the RRT and eligible study subgroup values were 132.56 and 16.51 respectively, compared to the candidate sample values which were 128.12 and 15.93.

Conversely, the study sample subgroup who did not meet CSE admission requirements demonstrated a significantly lower ability level in total test scores with much greater score variability, as compared to the RRT and eligible study subgroup and the candidate sample respectively. Descriptive statistics for the study subgroup not eligible for the CSE were 172.0 and 35.67 for the total IG score mean and standard deviation, and 87.67 and 35.72 for the total DM score mean and standard deviation.

Consequently, it was determined that the statistically significant differences in CSE test performance,
resulting from the initial MANOVAs and follow-up ANOVAs for the candidate and study samples, were influenced by composition characteristics of the study sample which contributed to lower test score means and quite high test score variances for this group. That is, the study sample subjects who were not eligible for the CSE tended to score lower than the RRT and eligible study subjects, as well as the candidate sample. Indeed, the significantly lower test scores observed for the study sample subgroup not eligible for the CSE can reasonably be considered evidence of instrument validity in that all the subjects in the candidate sample and those in the study sample who already held the RRT credential or met CSE admission requirements, performed at a significantly higher level. This is as expected if the examination is providing a valid assessment of the clinical abilities of the advanced level practitioner.

Although the statistically significant differences resulting from the initial MANOVAs and follow-up ANOVAs for the candidate and study samples clearly are influenced by the 106 subjects in the study sample who were not eligible for the CSE and tended to score lower, the candidate and study samples nevertheless represent examinee groups of observed different ability levels and different practitioner populations. That is, the candidate sample is representative of that somewhat
homogeneous subset of the respiratory care practitioner population who meet education and experience requirements for the CSE; the study sample is representative of the general population of all respiratory care practitioners employed in hospitals in the U.S., some of whom currently hold the RRT credential or are eligible for the CSE and some whom lack either the education or experience requirements and are therefore not eligible. The observed differences in ability levels for these two samples, who are considered representative of different examinee populations, provides a desirable condition for generalizability of the part one results in that the remarkably similar factor structures resulting for each sample indicates that the one factor solution is generalizable across examinee groups of different ability levels.

**The Dimensionality of CSE Problem Scores**

The second objective for part one of the study was to determine the dimensionality of the twenty problem score variables derived from an administration of the CSE to a large sample of NBRC candidates; the third objective was to determine the dimensionality for the twenty problem score variables derived from an administration of the same CSE test form to a second sample who completed the test solely for research purposes. This
allowed for the comparison of results across samples on the same test form to assess the generalizability of the underlying factor structure. Confirmatory factor analysis was deemed the dimensionality procedure of choice as previous factor analyses of physician PMP tests had demonstrated consistent factor solutions consisting of an information gathering and decision making dimension respectively, and the CSE was designed according to this two factor model with problem and total test scores derived to represent these dimensions.

Intercorrelation matrices for each of the samples were first generated to serve as input to the LISREL confirmatory factor analysis program. Inspection of the matrices revealed that all intercorrelations for the candidate data set, and nearly all for the study data set, were positive and statistically significant at the p< .01 level; this indicated that there was conceivably some association among the variables in each respective data set. Few of the intercorrelations for the candidate sample data set were greater than .30, probably due to the homogeneity of this group; conversely, nearly half of the intercorrelations for the study sample data set were greater than .30, probably due to the greater score variability of this group (see Appendix F).

As the LISREL procedure and determination of the rho goodness-of-fit measure were described in detail in
Chapter Four, it is sufficient to report here that both the candidate and study data sets failed to achieve an acceptable goodness-of-fit to the specified two factor information gathering (IG) and decision making (DM) model. The two intercorrelation matrices were therefore prepared for submission to the SAS principal axis exploratory factor analysis program, but it is important to note that the possibility of obtaining a two factor structure had not been ruled out at this point for many reasons. First, in specifying the model for confirmatory factor analysis, one specifies some parameters as zero for the target population factor loading matrix; according to the hypothesized model in this application, the IG problem score variables were specified to have zero loadings on the factor designated as decision making and conversely, the DM variables were specified to have zero loadings on the factor designated as information gathering. Consequently, the LISREL analysis could result in the IG and DM variables loading appropriately on their respective factors, but because of non-zero loadings for some of these variables on the opposite factor, an inadequate goodness-of-fit measure might result. Also, the LISREL program utilizes a maximum likelihood procedure for analysis which differs considerably from the principal axis factor method, and the differences in methods could produce different results.
Lastly, the exploratory analysis could result in a two factor structure, although the variables comprising each may have been mixed. That is, a two factor structure might be possible with a number of IG and DM score variables defining one factor perhaps based on similar problem types such as context of care, and another factor consisting of IG and DM variables from problems with some other common element.

However, results of the principal axis exploratory analyses demonstrated a clear one factor solution for each data set which accounted for approximately 80-86% of the variance in each case. The factor loadings in each solution were positively correlated with the factor and each variable loading for the study sample was slightly higher than its counterpart in the candidate solution. It appears that the greater variation in problem scores for the study sample resulted in higher correlations among these variables, and subsequently slightly higher factor loadings. Considering the factor loadings in each solution as representing the correlation between each variable and its respective factor, all were statistically significant at the p< .01 level, but four of the loadings in the candidate solution were very low (i.e., less than 0.30). Nevertheless, the statistically significant correlations with the factor for each solution indicate that the factor loadings are
not simply "chance" results; that is, each does significantly differ from zero.

Additionally, arranging the factor loadings for each solution in a highest-to-lowest order revealed that the loading order of the score variables was somewhat different in each case, but that the vast majority of the IG score variables loaded higher than the DM score variables on the factor for each solution. These consistent one factor solutions convincingly demonstrated that the underlying dimensionality of the twenty problem score variables was generalizable across groups of examinees who performed at different ability levels on the same CSE test form. Also, considering the very large sample size of the candidate group (n=2821) and the much smaller sample size of the study group (n=312), but acceptable in terms of subject-to-variable ratio, these consistent factor analysis results clearly demonstrated that the one factor dimensionality of the twenty problem score variables was quite stable.

Near the completion of these analyses, the CSE results for another sample of candidates, who had recently completed the December 1984 test form (comprised of ten different PMPs), became available. This December candidate sample consisted of 2546 subjects, similar in size and composition of first time and repeater candidates to the June 1984 candidate sample, and the
proportions of candidates passing/failing these two examinations was within two percent of one another. A matrix of intercorrelations for the twenty problem score variables for the new December candidate sample was obtained and submitted to the SAS principal axis factor analysis program with the same specifications as the prior two factor analyses. A consistent one factor solution resulted which accounted for 78.6% of the variance; also all factor loadings were positively correlated with the factor, statistically significant at the p < .01 level, and similar in magnitude to the June candidate sample factor loadings with only three loadings less than 0.30. Additionally, the majority of the IG variables loaded highest on the factor, and the overall results clearly demonstrated that the one factor solution was further generalizable across examinee groups of similar ability level and different test forms.

Lastly, Cronbach alpha reliability estimates for each of the three examinee groups demonstrated that the reliability of the twenty problem score variables comprising the factor in each solution were acceptable for such a criterion-referenced test. The estimates were 0.78 and 0.81 for the June and December candidate samples respectively, and 0.88 for the study sample.
Conclusions for Part One of the Study

Consistent factor solutions resulted for all three available examinee group data sets, demonstrating generalizability and stability of the one factor structure across different groups of examinees and different test forms. The factor solutions for the June 1984 CSE candidate sample and the study sample demonstrate that the problem score factor structure is stable and generalizable across different groups of examinees, who perform at statistically significant different ability levels and represent two distinct populations, but complete the same test form. The factor solution obtained for the December 1984 CSE candidate sample further demonstrates the one factor solution to be generalizable and stable across similar groups of examinees who complete different CSE test forms. These one factor results do not support the case specificity theory of medical problem solving as proposed by Elstein et al. (1978), because if this were so, one would expect a solution consisting of many factors with each factor defined by the IG and DM scores for a given problem or by problem scores of similar content. Nor do these results support the two factor solutions demonstrated by Juul et al. (1979) on measures derived from tests consisting of more than twenty branching PMPs, or by Skakun (1978) for measures derived from a test consisting of five linear PMPs.
Neither do the results of this study support those of Donnelly et al. (1974), who conducted individual factor analyses on content area scores derived from each of ten physician PMPs and an average across all ten; Donnelly et al. (1974) demonstrated a consistent two-factor solution for eight of the ten problem analyses as well as for the average score across problem analysis.

It is interesting to note that Donnelly et al. (1974) were the first authors to suggest that two scores representing information gathering and decision making ability respectively, were required to adequately describe PMP performance. However in the same article, these authors advanced a different conclusion based on subsequent analyses which are only briefly discussed. When these researchers calculated the Cronbach alpha reliability estimate for the scores achieved within each of the seven content areas across all problems, they discovered that the content areas which had clearly represented the information gathering factor (history, physical examination and laboratory studies) had coefficients of 0.75 to 0.87, but that the content area scores which represented the decision making factor (diagnosis, management and pathway) had considerably lower reliability estimates ranging from 0.19 to 0.44. The authors also determined that canonical correlations between the variables comprising each of the two factors for eight
of the ten problems were quite low (i.e., less than 0.30), and not statistically significant.

Donnelly et al. (1974) found this lack of correlation between the variables comprising each factor of interest and upon inspecting subject data, discovered that some subjects were able to render correct decisions with very little information while other subjects obtained most of the relevant information but were unable to render correct decisions. The authors concluded that the high reliabilities of the IG variables indicated that information gathering is a stable, more general ability which does not appear to vary appreciably from problem to problem. On the other hand, the low reliabilities of the DM variables were felt to indicate that decision making varies from problem to problem and therefore, that this ability is probably medical content or case specific. To further investigate this, Donnelly et al. (1974) conducted a factor analysis of the problem content area scores for all ten PMPs (total of seventy scores), and found that all but a few of the IG variables loaded on the first factor but that there was no general factor related to decision making. Consequently, Donnelly et al. (1974) failed to demonstrate a clear two factor structure for content area problem scores derived from all ten PMPs, despite their previous suggestion that two scores were required to adequately
describe performance.

There are some similarities between the Donnelly et al. (1974) analysis described above and those involved in part one of this study. For instance, although the specific procedures used by Donnelly et al. (1974) for the factor analysis of problem scores derived from the ten branching PMPs are not described by the authors, the use of intercorrelations for problem scores derived from an aggregate of branching PMPs is very similar to the data source method used for factor analysis in this study, and the results also failed to demonstrate a two factor structure. Additionally, although Donnelly et al. (1974) determined that the canonical correlation between the variables comprising each of the factors for the two factor results of individual problem scores to be low and insignificant, a later study by Juul et al. (1979), which demonstrated stable two factor solutions across examinee groups and test forms consisting of more than twenty branching PMPs, found that the factors were correlated in the range of 0.42 to 0.53 and that these correlations were statistically significant.

The generalizable and stable one factor results of part one of this study demonstrate the presumed underlying processes of information gathering and decision making not only to be highly related, but in fact to constitute one continuum, which may be perceived in a
similar fashion to Bloom's taxonomy of cognitive processes. That is to say, that the ability to select appropriate information options from a list may constitute a lower order recognition task, whereas the ability to synthesize the information obtained and render decisions may require complex higher order cognitive abilities. Moreover, it seems reasonable that practitioners may be able to select appropriate information to gather for a problem, but lack the ability to analyze the data in order to render effective decisions for patient care.

If one considers the information processing theory of human problem solving described by Simon and Newell (1971), the one factor results indicating a continuum of cognitive abilities with the IG variables generally more highly correlated with the factor in each case than the DM variables, may be further explicated. For instance, perhaps the ability to recognize appropriate information to select is one of the few gross characteristics of the human information processing system referred to by Simon and Newell (1971) which is relatively invariant over tasks and problem solvers. This would be consistent with the findings of Donnelly et al. (1974) that information gathering is a more stable and general ability, perhaps because it is an easier task to master. As the reader may recall, many studies have demonstrated that more information is gathered by practitioners when they
are allowed to select from a list of options as opposed to when they are required to generate their own alternatives (Blumberg, 1981; Goran, Williamson and Gonella, 1973; McCarthy, 1966; Newble, Hoare and Baxter, 1982; Norman and Feightner, 1981; Page and Fielding, 1980), which indicates that simple recognition influences performance.

Also, Simon and Newell (1971) described the problem solver as representing the task environment as a problem space with the structure of this space determining the information to be used for problem solving. Perhaps practitioners may generally be able to gather appropriate information but are not as generally able to efficiently organize the data into a problem space structure which is prerequisite to effective problem solving ability. This can be perceived as consistent with the hierarchical differences in cognitive ability represented by lower order cognitive processes such as knowledge and comprehension, as opposed to higher order cognitive processes such as analysis and synthesis. It is also possible that information gathering ability is more greatly influenced by the cueing effect demonstrated in many studies for structured response PMPs. Indeed, McCarthy (1966) concluded that such visual cueing in structured response PMPs acts as an aid to recall of relevant information; this would further contribute to
perceiving the stable and generalizable one factor structure demonstrated in this study as a continuum of possibly hierarchical cognitive processes in which information gathering ability is a necessary but insufficient condition for effective problem solving.

In conclusion, results of part one of this study neither support the case specificity theory of medical problem solving or the two factor model demonstrated by others. It seems reasonable that substantive interpretation of the data analyses for this part of the study may be perceived within the framework of Bloom's taxonomy of cognitive processes and the information processing theory of human problem solving described by Simon and Newell (1971). Nevertheless, the remarkably consistent one factor solutions resulting from part one of this study, which demonstrate the factor structure to be generalizable and stable across different samples of examinees and different CSE test forms, are at variance with all previous PMP dimensionality studies reported in the literature. It should be noted that although the previously discussed studies of PMP dimensionality which resulted in two factor solutions (Donnelly et al., 1974; Juul et al., 1979; Skakun, 1978) did utilize somewhat different derived scores and factor analysis methods, they also used "teacher-made" PMP instruments and considerably smaller samples of subjects (i.e., two
hundred or less) which were comprised exclusively of medical students enrolled at an educational institution. It could be argued that if the study of problem solving behavior is to be appropriately perceived as a person-environment interaction, then consideration for characteristics of the test and the subject's response system must be incorporated in the research design, and it is possible that unidimensional results might be obtained if the PMP instrument used was standardized to sufficiently represent the clinical domain for assessment and the samples for study of physician problem solving ability were larger, nationally representative and comprised of practicing physicians rather than medical students in training. On the other hand, it may be that the focus of clinical problem solving tasks is different for physicians (i.e., diagnostic emphasis) and PMP instruments therefore hold different properties for this clinical domain, and/or that the physician practice domain is just too broad and clinical responsibilities too all encompassing to allow for the discrete delineation and representative sampling of patient situations for general medical practice. Further discussion of the results of part one of this study, as well as implications for future research on the measurement properties of PMPs, is reserved for the chapter summary.
Part Two Discussion and Conclusions

Part Two of the study was designed to determine the relationship of measures of work experience derived from the Respiratory Therapy Work Experience Inventory (WEI), to CSE test performance. Only data on those 112 study subjects who returned a completed WEI (the respondent group), and their respective standardized factor score derived from the part one analysis, were used for this part of the study. As the reader may recall, this respondent group was a subset of the study sample; relative to that sample, the respondent group consisted of different proportions of subjects in terms of geographic representativeness but was very similar in terms of sex, age, and NBRC credential status proportions as well as mean years of experience.

Although the interpretation of results, especially as regards generalizability of the findings, must be tempered by the possibility of response bias due to the 36% return rate, differences in geographic representativeness and the slightly higher proportion of supervisor job title subjects, the respondent group was considered to constitute a reasonably representative subgroup of the total study sample in terms of demographic characteristics (see Table 3). Additionally, representativeness of the respondent group in terms of test
performance was also considered quite good as the mean CSE factor score was only slightly higher and the standard deviation slightly lower for this group as compared to the study sample standardized descriptive statistics (see Table 4). Therefore, the respondent group was considered reasonably well representative of the study sample in terms of demographic and test performance characteristics, thereby allowing the researcher to consider the respondent group representative of the population of practitioners comprising the respiratory therapy labor force in the U.S.

The first objective for part two of the study was to determine if statistically significant differences in CSE performance existed between subgroups of respondents categorized as currently holding different job levels. Results of the one-way ANOVA by job level, for the two subgroups categorized as currently holding an entry or advanced level job, demonstrated statistically significant differences in CSE performance (F test, p< .01). The direction of this difference was that the advanced group achieved considerably higher on the test score variable than the entry group.

As the reader may recall, jobs are categorized by the ratings provided in the Job Factors section of the WEI which consists of nine weighted items considered important for pay and classification of hospital
positions. Such things as prior education and training, job complexity, consequences of errors, and the degree to which the job involves direction of others are rated (see Appendix D for a copy of the WEI instrument). Considering the content of this WEI section, subjects categorized as advanced presumably hold more responsible positions such as therapist or supervisor; such hospital positions commonly require practitioners to have more education and experience as well as hold the RRT credential. In comparison, entry level positions in hospitals usually involve less responsibility and do not require advanced technical training, attainment of credentials or significant amounts of experience. Therefore, the advanced group might reasonably be expected to achieve higher CSE scores due to more extensive education and experience backgrounds, as was demonstrated by this analysis.

The second objective for part two of the study was to determine the statistical relationship between three quantified WEI measures, each representing a different aspect of clinical work experience, and CSE performance. A continuous score variable was derived for each of these three sections of the WEI (i.e., Work Situations, Job Responsibilities, and Knowledges, Skills and Abilities), which were collectively perceived as defining the respiratory therapy clinical domain in a similar
manner as the CSE (see Figures 1 and 2). Zero-order correlations for each section score with the CSE factor score were positive and statistically significant at the p<.01 level; however, those for Job Responsibilities and Knowledges, Skills and Abilities were slightly less than 0.30, while the Work Situations correlation was a little higher at 0.37. Because all three WEI scores were so highly intercorrelated (i.e., 0.64 to 0.73), only the Work Situations variable entered the multiple regression equation. Therefore, the results were identical to the zero-order correlation between this WEI variable and the CSE factor score, accounting for only 13.4% of the observed test score variance.

The high degree of multicollinearity between the WEI section scores was not expected because Nettles (1984) had reported an item rating intercorrelational analysis which revealed few values greater than 0.50; Nettles (1984) therefore concluded that there was little overlap in WEI items and that each WEI section was contributing somewhat unique information. However, results of this study demonstrate that WEI scores derived by section share a considerable amount of common variance, to the extent that insignificant prediction is provided by adding a second or third section score variable to the multiple regression equation. Moreover, although the zero-order correlation between each of the three WEI
section scores and the CSE factor score is positive, even the highest of these accounts for a very small proportion of variance in the test score variable. The extent to which this is so due to inadequacies of the WEI items, possible response bias, lack of integrity of the self-ratings, and/or the nature of the derived CSE factor score serving as the dependent variable, is unclear.

On the other hand, results of these correlational analyses are noted to be at variance with those of Farrand et al. (1982), Williamson (1965) and Kacmarek (1984), who found that performance on PMP tests was negatively correlated to years of experience. As the reader may recall, it was felt by this author that years of experience may be a misleading predictor because it reflects only quantity (as opposed to any indication of the quality) of experience. To investigate this belief, the zero-order correlation between CSE factor scores and the number of years of clinical experience reported by the respondent group on the WEI, was determined. A very low positive correlation (0.12), statistically not significant at the p< .05 level, resulted. This insignificant positive correlation is noted to differ from the statistically significant negative results referred to above and indicates that years of experience is less related to PMP test performance than quantified measures which reflect the nature of work experience.
In conclusion, the results of the analyses using the derived CSE factor score as the dependent variable, and the WEI Job Level category as well as the three WEI section scores as independent variables, support the contention of many researchers (DeGroot, 1965; Elstein, Shulman and Sprafka, 1978; McGuire, 1976; Simon and Newell, 1971), that experience in the substantive domain for which problems are to be solved is related to performance. Moreover, statistically significant differences in test performance did result as a function of job level (entry versus advanced), and although the most highly correlated WEI predictor failed to account for a very large proportion of variance in derived test scores, all three WEI predictors were positively and significantly correlated with the CSE factor score for this representative group of respiratory care practitioners.

Summary and Conclusions of the Study

It has been said that clinical competence is a complex phenomenon and not amenable to simple solutions in terms of measurement (Newble, 1979). Clinical competence consists of unique combinations of knowledges, skills and attitudes which are necessary for safe and effective practice in a given profession, and such competencies are noted to vary from simple behavioral
responses with few conditional alternatives to complex integrations of unlimited scope (Scanlon, 1978). It is generally agreed that no single test method is capable of providing a reliable and valid assessment of all the components which constitute competence (Marshall, 1977; McGuire, 1963, 1969; Newble 1979; Small, 1975). Consequently, it has become common in health practitioner credentialing to utilize multiple-choice examinations which generally provide reliable and valid assessments of practitioner knowledge but are somewhat insufficient indicators of clinical performance abilities, and to use simulation-based assessments to measure behaviors more closely associated with effective clinical performance.

Written simulation instruments consisting of an aggregate of representative patient management problems (PMPs) have been proposed as a useful assessment method to measure what has been termed clinical problem solving ability in patient management. Due to the differences between such instruments and typical objective cognitive knowledge tests, measures derived from written simulations are not amenable to traditional methods for determining the reliability and validity of psychological tests. Consequently, there has been much controversy regarding the psychometric properties of PMP tests based on a considerable amount of conflicting evidence.
Swanson (1984) stated that it is not surprising that results have varied due to the many differences in the instruments, derived measures and analysis procedures used for research in this area. Vu (1979) has further argued that a systematic approach in considering the different properties of each simulation instrument is necessary, although this recognizably limits the generalizability of research findings. To overcome the shortcomings of many previous studies, this research utilized the test results of large samples of subjects on a nationally standardized written simulation examination, which consisted of an aggregate of ten branching PMPs determined to represent the clinical domain for assessment in a well defined manner. Specifically, this research effort was directed at investigating the psychometric properties of the NBRC Clinical Simulation Examination with regard to opposing theories of medical problem solving presented in the literature and to the controversy regarding the relationship of clinical experience to performance on such tests.

The results of part one of this study failed to support the case specificity theory, or the two factor (i.e., information gathering and decision making) process model, of medical problem solving. However, the factor analyses resulted in remarkably similar one factor solutions for two groups of examinees who
demonstrated statistically significant differences in ability level on the same test form, and for another group of examinees who completed a different test form. These results clearly demonstrated the generalizability of the consistent one factor solution across different samples of subjects and different aggregates of problems comprising a test form. It seems clear that regardless of the ability level of the examinee group, or the group of problems selected according to the established decision rules to comprise a CSE test form, the underlying dimensionality of the twenty problem scores is stable and singular (i.e., unidimensional).

The consistent one factor structure which resulted for each of the three examinee samples demonstrated that the cognitive processes underlying performance on this PMP test, which is carefully designed to assess clinical abilities in the relatively discrete domain of respiratory therapy, exist on a singular continuum, despite the fact that the CSE was designed according to the two dimensional model and utilizes two different item types presumed to elicit distinctly different behaviors or cognitive processes. Noting that the IG variables were generally more highly correlated with the factor and considering the information processing theory of human problem solving as described by Simon and Newell (1971), it was substantively concluded that the ability to
select appropriate information is a necessary but insufficient condition for effective decision making in patient management. It is notable that Donnelly et al. (1974) had arrived at a similar conclusion based on analyses of problem score measures derived from a physician test consisting of ten branching PMPs.

The question of what to term the one factor resulting from each analysis is not easily answered. It could be argued that this represents a case of what Spearman called the "g" or general skill/ability factor. It is possible that this factor represents a general knowledge factor for the respiratory therapy clinical practice domain. However, the author prefers to term this factor "respiratory therapy clinical care ability" to underscore the differences between the nature of the multiple choice assessment format and that of the written simulation branching patient management problem format, but suggests that future research should investigate the relationship of scores achieved on the advanced level multiple-choice examination to the derived CSE factor score in order to further explicate the meaning of this factor. Such further research is necessary to provide evidence as to whether the CSE and similar PMP tests provide measures of the somewhat ill-defined cognitive processes presumed to underlie test performance and constitute clinical problem solving ability.
Another question which remains unresolved by this research is what summary score method is the most appropriate to describe CSE performance. The stable and generalizable one factor structure resulting from these analyses indicate that a derived singular factor score is appropriate. For instance, the precise order and magnitude of the problem score loadings could be determined following each CSE administration and a total test minimum pass level factor score derived by utilizing the minimum pass level score for each of the problem score variables. In order to provide insight as to the relative merits the singular factor score method indicated by this research and the two score method currently used by the NBRC, it would be appropriate to undertake a series of classification studies on multiple samples. Considering that the CSE is a criterion-referenced examination designed to identify two groups of examinees (pass versus fail), such a series of classification analyses would allow for comparison of the different summary score methods in terms of assignment to group, and the related "misclassification" of examinees according to different methods.

Lastly, results of the work experience analyses indicate that quantified measures of work experience are significantly related to PMP test performance in the same clinical domain. The logical and meaningful result
of statistically significant differences in test performance by job level, revealed that practitioners with more extensive education and experience backgrounds do perform better. Although the other three WEI measures were shown to be positively and significantly correlated with the CSE factor score, they failed to account for very much variance in these test scores. To what extent this result may have been affected by inadequacies of the WEI items, lack of integrity of the self-ratings, and/or the nature of the derived CSE factor score, is unclear. Replication of this part of the study, utilizing a larger, perhaps randomly selected sample may help to clarify the relationship between WEI measures of work experience and the derived CSE factor scores.

In conclusion, it is hoped that the results of this study will contribute to the body of research regarding the psychometric properties of PMP tests, including the nature of cognitive processes hypothesized to underlie such test performance and the relationship of clinical experience in the domain for which problems are to be solved. However, the stable and generalizable one factor structure demonstrated by this research is noted to be unique in that no such findings have previously been reported in the literature. Consequently, the author agrees with the conclusions of Marshall (1983), and Page and Fielding (1980), that the lack of
conclusive evidence regarding the validity of PMP tests in assessing that set of complex and poorly understood processes required for effective clinical performance, may be a problem of the state of the art of measurement in this area and that more research is needed before tests of this type are discarded for what may be the wrong reasons (i.e., they do not fit the mathematical model for classical test theory very well). Indeed, results of this study indicate that clearly defining the clinical domain for assessment and utilizing systematic procedures in test development, result in a stable unidimensional test. However, much more research is needed pertaining to the nature of clinical problem solving ability and its measurement by PMP tests, as well as the nature of the perplexing and as yet unclearly defined cognitive processes involved in human problem solving, before definitive conclusions regarding the psychometric properties of PMP tests and their appropriate role in health practitioner clinical assessment may be rendered.
BIBLIOGRAPHY


APPENDIX A: Sample CSE Score Report
NATIONAL BOARD FOR RESPIRATORY CARE

CLINICAL SIMULATION EXAMINATION FOR RESPIRATORY THERAPISTS

SEP 24, 1984

ETS FILE NUMBER: SOCIAL SECURITY NUMBER:

---

YOUR RESULT: PASS

*** SCORES BY SIMULATION PROBLEM ***

<table>
<thead>
<tr>
<th>SIMULATION PROBLEM</th>
<th>INFORMATION GATHERING SECTIONS</th>
<th>DECISION MAKING SECTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YOUR</td>
<td>PASSING</td>
</tr>
<tr>
<td>1. RAY MURPHY</td>
<td>45</td>
<td>37</td>
</tr>
<tr>
<td>2. ANTHONY SANTORO</td>
<td>38</td>
<td>24</td>
</tr>
<tr>
<td>3. MARY LUDWIG</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>4. DENNIS MILLER</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>5. DIANE ROSE</td>
<td>14</td>
<td>11</td>
</tr>
<tr>
<td>6. CARLTON JONES</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>7. JERRY GORDON</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>8. STEPHEN SANTIAGO</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>9. HAROLD COHEN</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>10. KIM CHAN</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>TOTAL SCORES (RAW):</td>
<td>239</td>
<td>191</td>
</tr>
<tr>
<td>TOTAL SCORES (%)</td>
<td>89%</td>
<td>71%</td>
</tr>
</tbody>
</table>

SUCCESSFUL COMPLETION OF THE CLINICAL SIMULATION EXAMINATION ALONE DOES NOT REFLECT HRT STATUS: SUCCESSFUL COMPLETION OF THE WRITTEN REGISTRY EXAMINATION IS ALSO REQUIRED.

*** CONTENT AREA ANALYSIS ***

<table>
<thead>
<tr>
<th>CONTENT CATEGORY</th>
<th>INFORMATION GATHERING</th>
<th>DECISION MAKING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCORE</td>
<td>MAX</td>
</tr>
<tr>
<td>1. AIRWAY CARE/MAINTENANCE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2. BRONCHIAL HYGIENE TECHNIQUES</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3. CARDIOPULMONARY RESUSCITATION</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>4. CARDIO-RESPIRATORY MONITORING</td>
<td>75</td>
<td>87</td>
</tr>
<tr>
<td>5. CARDIO-RESPIRATORY REHAB.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. CLINICAL ASSESSMENT</td>
<td>98</td>
<td>103</td>
</tr>
<tr>
<td>7. CONT. VENTILATORY SUPPORT TECH.</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8. INFECTION CONTROL</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9. LABORATORY STUDIES ASSESSMENT</td>
<td>42</td>
<td>50</td>
</tr>
<tr>
<td>10. OXYGEN THERAPY</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11. PATHOPHYSIOLOGY</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12. PHARMACOLOGY</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13. PULMONARY FUNCTION STUDIES</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>TOTAL SCORES (RAW):</td>
<td>239</td>
<td>269</td>
</tr>
</tbody>
</table>

SUCCESSFUL COMPLETION OF THE CLINICAL SIMULATION EXAMINATION ALONE DOES NOT REFLECT HRT STATUS: SUCCESSFUL COMPLETION OF THE WRITTEN REGISTRY EXAMINATION IS ALSO REQUIRED.
APPENDIX B: CSE Test Matrix - June 1984 Test Form
<table>
<thead>
<tr>
<th>CONTENT AREAS/PROCESS</th>
<th>INFO. GATHERING</th>
<th>TOTAL</th>
<th>C</th>
<th>C&lt;sup&gt;1&lt;/sup&gt;</th>
<th>DECISION MAKING</th>
<th>TOTAL</th>
<th>C</th>
<th>C&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Airway Care/Maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2,1,5,1,3,(2)</td>
<td>14</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>2. Bronchial Hygiene Techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2,2,(1),2,4,(2),1</td>
<td>14</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>3. Cardiopulmonary Resuscitation</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1,1,3,4,(1)</td>
<td>10</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>4. Cardio-Respiratory Monitoring</td>
<td>2,3,2</td>
<td></td>
<td>7</td>
<td>7</td>
<td>1,1,1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5. Cardio-Respiratory Rehabilitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6. Clinical Assessment</td>
<td>1,1,1,1,1,1,1,2</td>
<td></td>
<td>9</td>
<td>9</td>
<td>1,1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>8. Infection Control</td>
<td>1,2</td>
<td></td>
<td></td>
<td></td>
<td>1,2</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>9. Laboratory Studies Assessment</td>
<td>1,(1),1,1,1,1,1</td>
<td></td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1,1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10. Oxygen Therapy</td>
<td>1,(1),1,1,1,1,1,1</td>
<td></td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Pathophysiology</td>
<td>1,1</td>
<td></td>
<td></td>
<td></td>
<td>1,1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>12. Pharmacology</td>
<td>2,(1),2,2,1</td>
<td></td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1,1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13. Pulmonary Function Studies</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1,1,1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

**Total Sections:** Critical Path IG 25 DM 73  
Entire Exam IG 26 DM 92
APPENDIX C: CSE Sample Patient Management Problem

Note: Responses in boxes are exposed on the sample problem for illustrative purposes; for actual testing, they would be concealed by latent print.
SIMULATION 2: Carl Rivers—A Ventilator Patient

You are the respiratory therapist assigned to a 4-bed ICU on the midnight shift. On entering the unit, you find that 52-year-old Carl Rivers is being mechanically ventilated with a volume ventilator. Mr. Rivers is being mechanically ventilated with IMV and appears to be resting comfortably. He is 182 cm (6 feet) tall and weighs 77 kg (170 lbs.).

NOW GO TO SECTION A

SECTION A—Carl Rivers

During your initial ventilator check, which of the following would you evaluate?

(Select as many as you consider indicated in this section.)

| A-1. Exhaled tidal volume | A-1. 900ml** |
| A-3. Ventilator flow rate | A-3. 50 liters/min.** |
| A-4. Ventilator mandatory rate | A-4. 8/min.** |
| A-7. FIO₂ | A-7. 0.4 ** |
| A-8. Effective static (plateau) compliance | A-8. 51ml/cm H₂O.** |
| A-10. Patient temperature | A-10. 37.8°C.** |
| A-11. Results of most recent ABG analysis | A-11. PaO₂ 78 torr, pH 7.39, PaCO₂ 42 torr, HCO₃ 25 mEq/liter, SaO₂ 94%.** |
| A-13. Heart rate | A-13. 74/min.** |
SECTION A—Continued

A-15. Blood pressure
A-16. Test of spirometer alarm
A-17. Response to painful stimuli
A-18. Pressure limit
A-19. Character of airway secretions
A-20. Gas temperature
A-21. FEV₁₃
A-22. Water level in humidifier

A-15. 114/68 mmHg.**
A-16. Functioning.**
A-17. Noted.**
A-18. Set at 40 cm H₂O.**
A-19. Appear to be mucoid, tenacious and very abundant.**
A-20. 35°C.**
A-21. Unable to obtain.**
A-22. Filled to maximum level.**

WHEN YOU HAVE COMPLETED YOUR SELECTIONS IN THIS SECTION, DEVELOP RESPONSE A-23.
A-23. Go to Section C.**

SECTION B—Carl Rivers

Two hours later, auscultation of the patient's chest reveals wheezing bilaterally. The patient appears to be in no significant respiratory distress. You would recommend the administration of which of the following? (Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

B-1. An in-line ultrasonic treatment
B-2. 0.5 ml. isoetharine (Bronkosol) and 3 ml. normal saline via a side-arm nebulizer
B-3. A subcutaneous injection of epinephrine
B-4. Cromolyn sodium (Intal) via aerosol
B-5. 1 ml. isoproterenol (Isuprel 1:100) with 3 ml. normal saline via a side-arm nebulizer

B-1. Physician disagrees. Make another selection in this Section.**
B-2. Physician agrees. Done. The patient's wheezing subsides. Go to Section G.**
B-3. Physician disagrees. Make another selection in this Section.**
B-4. Physician disagrees. Make another selection in this Section.**
B-5. Physician disagrees. Make another selection in this Section.**
SECTION C—Carl Rivera

At 12:15 a.m., the pressure limit alarm on Mr. River's ventilator begins sounding. You immediately suction him and reconnect the ventilator circuit; however, the alarm continues to sound. Which of the following is the first action you would take?

(Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

C-1. Page a physician to get an order for a stat. arterial blood gas analysis.

C-1. Physician unavailable. Make another selection in this Section.**

C-2. Decrease the tidal volume and flow rate

C-2. Done. The pressure alarm is still sounding. Make another selection in this Section.**

C-3. Increase the pressure limit setting and record the change.

C-3. Done. The pressure alarm is still sounding. Go to Section E.**

C-4. Ventilate the patient with a resuscitator bag and auscultate his chest.

C-4. Done. Breath sounds are absent on the left side. Physician arrives and orders a stat. chest x-ray, which shows that the tube is in the right main stem bronchus. Go to Section D.**

C-5. Increase the pressure limit setting and tidal volume to insure adequate ventilation.

C-5. Done. The pressure alarm is still sounding. Go to Section E.**

SECTION D—Carl Rivera

What would you recommend now?

(Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

D-1. Withdraw the endotracheal tube 2 cm. and auscultate the patient's chest

D-1. Physician agrees. Done. Go to Section I.**

D-2. Deflate the cuff, withdraw the endotracheal tube 2 cm., inflate the cuff, and auscultate the patient's chest

D-2. Physician agrees. Done. Go to Section I.**

D-3. Increase the tidal volume

D-3. Physician disagrees. Make another selection in this Section.**

D-4. Deflate the cuff and auscultate the patient's chest.

D-4. Physician disagrees. Make another selection in this Section.**

D-5. Extubate the patient and reintubate with a smaller endotracheal tube

D-5. Physician disagrees. Make another selection in this Section.**
SECTION E—Carl Rivera

Now what would you do?
(Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

E-1. Ventilate the patient with a resuscitator bag and auscultate his chest.

E-1. Done. Breath sounds are absent on the left side. Physician arrives and orders a stat chest X-ray, which shows that the tube is in the right main stem bronchus.
Go to Section D.**

E-2. Add 5 ml. of air to the endotracheal tube cuff.

E-2. Done. The pressure alarm is still sounding. Make another selection in this Section.**

E-3. Sigh the patient repeatedly

E-3. Done. The pressure alarm is still sounding. Make another selection in this Section.**

E-4. Decrease the tidal volume and the flow rate.

E-4. Done. The pressure alarm is still sounding. Make another selection in this Section.**

SECTION G—Carl Rivera

At 6:15 a.m., the last ventilator check is made, all alarms are checked and found to be operable and sterile water is added to the humidifier. Immediately after this last check, you go to the conference room to chart. The nurse who has been taking care of Mr. Rivera comes into the conference room and informs you that the low volume alarm is sounding. Which of the following would you do now?
(Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

G-1. Inform the nurse that you must get your charting done and that you will be there in a minute.

G-1. The nurse insists that you do something now. Make another selection in this Section.**

G-2. Ask the nurse to recheck the patient's tidal volume with a respirometer.

G-2. The nurse refuses. Make another selection in this Section.**

G-3. Ask the nurse to increase the volume setting until a 900 exhaled tidal volume is obtained.

G-3. The nurse refuses. Make another selection in this Section.**

G-4. Go to the patient's bedside, auscultate his chest, and check for an obstruction of the endotracheal tube.

G-4. Done. Aeration is diminished bilaterally; however, the endotracheal tube is patent. The physician in the ICU asks what your next action would be.
Go to Section J.**

G-5. Go to the patient's bedside, auscultate his chest, and check for a leak in the ventilator system.

G-5. Done. Aeration is diminished bilaterally and you discover a leak in the ventilator circuit.
Go to Section H.**
SECTION H—Carl Rivers

Which of the following would you do now?
(Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

H-1. Leave the unit to obtain another ventilator.

H-2. Extubate the patient and ventilate with a resuscitator bag and mask.

H-3. Administer 60% O₂ via a Briggs' adapter (T-piece).

H-4. Ventilate the patient with a resuscitator bag via the endotracheal tube.

H-5. Increase the tidal volume to 1,000 ml.

End of the problem.

SECTION I—Carl Rivers

Mr. Rivers is now being mechanically ventilated and the system pressure has returned to normal level. Mr. River's breath sounds are equal bilaterally. Which of the following procedures and/or laboratory determinations would you recommend at this time?
(Select AS MANY as you consider indicated in this Section.)

I-1. Chest x-ray

I-2. C-T scan of head

I-3. Stat. serum electrolytes

I-4. Insertion of a flow-directed pulmonary artery catheter (Swan-Ganz)

I-5. Stat. arterial blood gas analysis

I-6. Sputum culture

I-7. CBC

I-8. Go to Section B.
SECTION J—Carl Rivers

What would you do at this time?
(Choose ONLY ONE unless you are directed to "Make another selection in this Section.")

<table>
<thead>
<tr>
<th>J-1.</th>
<th>Check for a leak in the ventilator system.</th>
<th>J-1.</th>
<th>Done. You discover a leak in the ventilator circuit. Go to Section H.**</th>
</tr>
</thead>
<tbody>
<tr>
<td>J-3.</td>
<td>Increase the tidal volume to 1,000 ml.</td>
<td>J-3.</td>
<td>Physician disagrees. Make another selection in this Section.**</td>
</tr>
<tr>
<td>J-4.</td>
<td>Add 6 ml. of air to the endotracheal tube cuff.</td>
<td>J-4.</td>
<td>Physician disagrees. Make another selection in this Section.**</td>
</tr>
</tbody>
</table>
APPENDIX D: Work Experience Inventory (WEI) Instrument
WORK EXPERIENCE INVENTORY FOR RESPIRATORY THERAPY PRACTITIONERS

The Respiratory Therapy Work Experience Inventory (WEI) was designed to enable respiratory therapy practitioners to describe their clinical and work experience in a systematic way, and to help practitioners better understand and demonstrate the experience they have gained in the respiratory therapy profession. The questions in the WEI are intended to help you describe the extent of your exposure to respiratory care activities and responsibilities during your work experience to date. The WEI is not a test with right or wrong answers; instead, it provides a standardized format for you to describe important aspects of your current and past work experience. The work experience you describe may have been acquired in full-time or part-time employment in the respiratory therapy profession.

The WEI was developed by Educational Testing Service in conjunction with the National Board for Respiratory Care and provides the basis for the NBRC's national job analysis for the Advanced Respiratory Therapy Practitioner. The WEI is divided into five parts:

I. A brief description of your employment history
II. A description of your work experience in terms of Job Factors
III. A description of your work experience in terms of Work Situations and Climates
IV. A sample of responsibilities that you might have performed
V. A list of knowledges, skills, and abilities that you might have acquired in the course of your work experience

Please carefully complete the following information to insure prompt receipt of your Work Experience Inventory report. You should receive your summary within eight weeks after we receive your completed inventory.

PLEASE PRINT

NAME:____________________________________________________

Last     First     M.I.

ADDRESS: ________________________________________________

CITY: ____________________________________________________

STATE: _______ ZIP CODE: ______________

TELEPHONE #: (____)
(where you can be reached between the hours of 9:00 am and 5:00 pm CST.)

Copyright © 1984 by Educational Testing Service. All rights reserved.
PART 1: 
EMPLOYMENT HISTORY

Please provide the requested information below for your past employment most significantly related to the practice of respiratory care. Describe your current or most recent job first, entering these data for Job A on this inventory. The inventory allows you to summarize your experience for the last three jobs you have held.

**JOB A**

**Job Title (circle one choice for items 1-7; circle two choices for item 8)**

1. Department Director
2. Assistant Department Director
3. Instructor
4. Rehabilitation Practitioner
5. Clinical Supervisor
6. Pulmonary Lab Supervisor
7. Pulmonary Lab Practitioner
8. Respiratory Care Practitioner:
   a. adults
   b. neonatal pediatric
   c. both

---

**Date of hire (MO/yr)**

**Date of termination (MO/yr)**

**Months of experience**

---

**JOB B**

**Job Title (circle one choice for items 1-7; circle two choices for item 8)**

1. Department Director
2. Assistant Department Director
3. Instructor
4. Rehabilitation Practitioner
5. Clinical Supervisor
6. Pulmonary Lab Supervisor
7. Pulmonary Lab Practitioner
8. Respiratory Care Practitioner:
   a. adults
   b. neonatal pediatric
   c. both

---

**Date of hire (MO/yr)**

**Date of termination (MO/yr)**

**Months of experience**

---

**JOB C**

**Job Title (circle one choice for items 1-7; circle two choices for item 8)**

1. Department Director
2. Assistant Department Director
3. Instructor
4. Rehabilitation Practitioner
5. Clinical Supervisor
6. Pulmonary Lab Supervisor
7. Pulmonary Lab Practitioner
8. Respiratory Care Practitioner:
   a. adults
   b. neonatal pediatric
   c. both

---

**Date of hire (MO/yr)**

**Date of termination (MO/yr)**

**Months of experience**
**PART II:**

**DESCRIPTION OF WORK EXPERIENCE IN TERMS OF JOB FACTORS**

Rating Scale - use the scales below to indicate the degree to which each job element or characteristic was present in each of your prior Jobs A, B, and C. Record your judgement in the column to the right, selecting the option which best describes each of your past jobs.

1. Prior Education and Training: The amount of academic or specialized technical education or training considered a prerequisite to performing the job.

   The minimum education required in my past positions in Respiratory Therapy was:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High school diploma or equivalent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>One year Respiratory Therapy program or equivalent on-the-job training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Two year Respiratory Therapy program or equivalent on-the-job training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Four year college degree Respiratory Therapy program or equivalent on-the-job training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Graduate degree or equivalent on-the-job training</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Prior Knowledge: Job-related knowledge beyond that acquired through the formal education and training listed above; generally gained through related on-the-job training or experience.

   Knowledge:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>position required limited knowledge of routine work procedures or methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>position required general knowledge of some technical procedures or activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>position required working knowledge of several technical procedures or work activity areas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>position required ability to teach or supervise several technical procedures or work activity areas</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Prior Experience: Job related skills beyond those acquired through the formal education and training listed above; generally gained through related on-the-job training or experience.

   Experience:

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no previous work experience required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>up to six months work experience required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>six months to one year work experience required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>one to two years work experience required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>two to four years work experience required</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>more than four years work experience required</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Attainment of Job Proficiency: The length of time required to become proficient in job skills after employment.

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>less than one month</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>more than one but less than three months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>more than three but less than six months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>more than six but less than twelve months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>one year or more</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. **Job Complexity**: The requirements for exercising independent judgement in making work-related decisions.

1 = very structured - required minor decision making in the application of prescribed methods or procedures
2 = generally structured - required judgement or originality in selecting or applying alternative methods or procedures
3 = only partially structured - judgement frequently required to solve problems or make decisions in areas not specifically covered by established practices and procedures
4 = essentially unstructured - activities governed only by broad objectives and policies; available guides or precedents are limited

<table>
<thead>
<tr>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
</table>

6. **Direction of Others**: The extent of the job's organizational responsibility.

1 = no formal supervisory responsibility; was occasionally involved in some responsibility for checking the work of others
2 = was assigned responsibility for helping to orient and train others; assigning and checking their work; occasionally acted as lead worker
3 = provided immediate supervision over one or more positions; formally planned, scheduled, directed, and coordinated work of others
4 = in addition to immediate supervision, assigned responsibility in recommending compensation, staff selection, disciplinary action, grievances, performance appraisal, and similar supervisory duties; perform limited technical work

<table>
<thead>
<tr>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
</table>

7. **Contacts**: The responsibility for contacts or relationships with other hospital employees, patients or their families, other medical staff, and the public at large.

1 = contacts generally with immediate associates and supervisors
2 = contacts with patients or others usually involve discussion or transmission of established, patterned information
3 = contacts with patients or others involve limited interpretation of information or formulation of responses based upon general knowledge of hospital policies, procedures, and programs
4 = contacts with patients or others involve exchange of important information or the dissemination of information requiring careful interpretation of hospital policies
5 = contacts with patients or others involve communication of highly complex, important, and highly confidential information

<table>
<thead>
<tr>
<th>Job A</th>
<th>Job B</th>
<th>Job C</th>
</tr>
</thead>
</table>
8. Consequences of Errors: The extent to which the job requires accuracy and thoroughness to avoid potential errors and the consequences of those errors. Consider daily activities and not extreme situations that rarely occur.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Errors cause little danger to safety of patient or other staff members</td>
</tr>
<tr>
<td>2</td>
<td>Errors require minor expenditure of time for correction and limited exposure to minor health and safety risks</td>
</tr>
<tr>
<td>3</td>
<td>Errors may cause moderate delays in services affecting other departments, personnel, or patients, or moderate health and safety risks</td>
</tr>
<tr>
<td>4</td>
<td>Errors may be costly, life-threatening or have serious effect on relations with physicians, employees, patients, or others</td>
</tr>
</tbody>
</table>

9. Physical Effort: The physical effort required to perform the job satisfactorily.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minimum of physical effort of a fatiguing nature</td>
</tr>
<tr>
<td>2</td>
<td>Small amount of physical effort in the lifting of lightweight objects; work performed in a physically comfortable position</td>
</tr>
<tr>
<td>3</td>
<td>Continuous physical activity required; continuous bending, stooping, walking, climbing stairs, or sitting in one position</td>
</tr>
<tr>
<td>4</td>
<td>Arduous physical work throughout the work shift</td>
</tr>
</tbody>
</table>
PART III:
WORK SITUATIONS AND CLIMATES

Listed below are a series of situations that you might have encountered in your total work experience. If you have not encountered the situation in your work experience, put a zero (does not apply) on the line preceding the item number. If the item does apply, decide how significant a part of the position(s) you held it represents, by allotting a value of 0 to 7 to the item using the scale below. Enter the value on the line preceding the item number.

0 = Not applicable
1 = A minor part of a position I held
2 = ......................................
3 = ......................................
4 = A most significant part of a position I held

1. Advising or instructing others (coworkers, patients) on techniques or procedures through demonstration
2. Instructing in continuing education or inservice training programs
3. Attending continuing education or inservice education programs
4. Managing the fiscal and personnel resources of the department
5. Developing policies to comply with legal requirements and professional ethics
6. Evaluating new clinical procedures and technologies
7. Work in adult intensive care unit
8. Work in pediatric/neonatal intensive care
9. Work in recovery room
10. Work in adult general floor care
11. Work in pediatric/neonatal general floor care
12. Work in emergency room
13. Work in pulmonary function lab (excluding blood gas analysis)
14. Work in operating room
15. Work in Blood Gas lab
16. Work in research lab
17. Work in other diagnostic lab (specify: ..................................................)
18. Work in outpatient areas
19. Work in home care rehabilitation
20. Clinical instructing in affiliation with a Respiratory Therapy program
21. Participate in emergency land/air transport
22. Equipment maintenance and repair
PART IV:
POSITION RESPONSIBILITIES

Listed below are a series of responsibilities of the Advanced Respiratory Therapy Practitioner. If the responsibility does not apply put a zero on the line preceding the number. If the responsibility does apply to your experience, decide how SIGNIFICANT it was to the position(s) you held and allot a value from 0 to 7 to the item using the scale below. Enter the value on the line preceding the item number.

0 = Not applicable
1 = A minor part of a position I held
2 = ........................................
3 = ........................................
4 = An important part of a position I held
5 = ........................................
6 = ........................................
7 = A most significant part of a position I held

NOTE: These tasks are to be viewed as always being performed within any constraints of the physician's order.

Review existing data in patient record:

23. patient history, physical examination and current vital signs, admission and current respiratory care orders.
24. pulmonary function values and blood gas results.
25. central venous pressure, pulmonary wedge pressure, pulmonary artery pressures, cardiac output.
26. electrocardiogram.

Collect and evaluate additional pertinent clinical information by:

27. inspection: breathing pattern, cyanosis.
28. auscultation: increased, decreased, absent, or unequal breath sounds.
29. auscultation: rhonchi or râles (crackles).

Inspect chest x-ray to determine:

30. position of endotracheal or tracheostomy tube.
31. presence of or changes in pneumothorax or subcutaneous emphysema, presence of or changes in consolidation and/or atelectasis.

Perform and evaluate bedside procedures:

32. Perform and evaluate bedside spirometry.
33. Interview patient to determine clinical status: level of consciousness, ability to cooperate, presence of dyspnea and/or orthopnea.

Recommend collection of additional pertinent data via lab test:

34. chest x-ray.
35. blood gas analysis.
36. spirometry before and/or after bronchodilator.
Evaluate the following test:

37. spirometry before and/or after bronchodilator.
38. blood gas analysis (PO₂, PCO₂, pH).

Determine appropriateness of prescribed therapy and goals:
39. for identified pathophysiological state.

Select, assemble, and check the following equipment for proper function, operation, and cleanliness:
40. oxygen administration devices.
41. humidifiers and aerosol generators.
42. resuscitation devices: manual resuscitators, demand valves.
43. ventilators.
44. artificial airways.
45. regulators and flow meters.
46. oxygen analyzers.
47. air compressors.
48. IMV, CPAP, and PEEP circuits.
49. IPPB and incentive breathing devices.
50. environmental devices: mist tents, incubators, etc.
51. respirometers.
52. hemodynamic monitoring devices: monitors, arterial lines, PA lines, etc.

Assure selected equipment cleanliness:
53. perform procedures for disinfection and sterilization.
54. monitor effectiveness of sterilization procedures.

Perform quality control procedures for:
55. blood gas analyzers.
56. pulmonary function equipment.

Explain planned therapy:
57. and goals to patient.

Conduct prescribed therapeutic procedures to achieve maintenance of a patent airway:
58. perform endotracheal intubation.
59. maintain proper cuff inflation and position of endotracheal or tracheostomy tube.
60. extubate the patient.
Conduct prescribed therapeutic procedures to achieve the removal of broncho-pulmonary secretions, and/or bronchodilation, and/or reduction of mucosal edema:

61. perform postural drainage.
62. suction endotracheal and tracheostomy tube.
63. administer bland aerosol therapy.
64. administer prescribed pharmacologic agents (bronchodilators, mucolytics, etc.).

Conduct prescribed therapeutic procedures to achieve adequate spontaneous and artificial ventilation:

65. instruct and monitor techniques of incentive spirometry.
66. initiate and adjust IPPB therapy.
67. initiate and adjust continuous mechanical ventilation when no settings are specified.
68. institute and modify weaning procedures.

Conduct prescribed therapeutic procedures to achieve adequate arterial and tissue oxygenation:

69. initiate and adjust O₂ therapy.
70. initiate and adjust CPAP and PEEP therapy.

Evaluate and monitor patient's response to respiratory therapy:

71. measure and record vital signs.
72. auscultate chest and record changes.
73. observe changes in sputum production.
74. interpret hemodynamic data: CVP, PAP, wedge pressure and/or cardiac output results.
75. determine compliance values.
76. observe patient adverse reaction and terminate treatment.

Modify treatment techniques on patient response:

77. to IPPB: adjust flow, volume, etc.
78. to incentive breathing devices: goals, equipment, etc.
79. to aerosol therapy: change dosage of medication, aerosol output, temperature, etc.
80. to O₂ therapy: adjust flow, adjust FIO₂, etc.
81. to chest physiotherapy: alter position, technique, etc.
82. to CMV: modify settings, change modes (IMV, CPAP, etc.), etc.
83. to CMV: institute weaning, change weaning procedures.

Initiate optimum emergency resuscitation:

84. perform CPR.
85. recommend defibrillation.
Initiate and conduct pulmonary rehabilitation and home care:

- 86. modify respiratory therapy procedures for use in home.
- 87. explain and instruct planned therapy goals to patient and patient's family.
- 88. monitor and maintain home respiratory therapy equipment.

Maintain records and communication:

- 89. Record therapy and results using conventional terminology as required by hospital policy and regulating agencies.
- 90. Communicate information regarding patient's clinical status to appropriate members of the health care team.
- 91. Communicate information relevant to coordinating patient care, e.g., scheduling, avoiding conflicts, sequencing of therapies.

Administer Respiratory Therapy department:

- 92. evaluate and purchase equipment.
- 93. develop policies and procedures: clinical and administrative.
- 94. develop quality assurance programs, perform quality assurance audits.

Supervise respiratory care staff and services:

- 95. guide staff development, evaluate staff performance.
- 96. discipline departmental personnel.
- 97. assess priorities for delivery of patient care.

Perform inservice training:

- 98. develop education and training programs.
- 99. conduct inservice training for departmental staff.
- 100. conduct inservice training for nondepartmental staff (physicians, nurses, public, etc.)

Perform formal/academic respiratory therapy education:

- 101. perform clinical instruction.
- 102. instruct in formal classroomconference setting.

Assist physician performing:

- 103. bronchoscopy.
- 104. invasive monitoring: central venous pressure, pulmonary artery catheters, arterial lines, etc.
PART V:  
RESPIRATORY CARE KNOWLEDGES, SKILLS, AND ABILITIES

The following list of knowledges, skills, and abilities may be significant for respiratory therapy practitioners. For each item, decide how significant each was in the performance of your previous related work experience, allotting a value from 0 to 7 using the SIGNIFICANCE scale below. Enter the value on the line preceding the item number.

0 = Not applicable
1 = A minor part of a position I held
2 = ....................................
3 = ....................................
4 = An important part of a position I held
5 = ....................................
6 = ....................................
7 = A most significant part of a position I held

Airway Care/Maintenance:
105. airway devices selection and operation
106. endotracheal and tracheostomy tube care
107. tracheobronchial suctioning
108. intubation/extubation assessment and techniques
109. emergency maneuvers

Bronchial Hygiene Techniques (Humidity/Aerosol Therapy, IPPB-Incentive Spirometry, Pulmonary Physiotherapy):
110. equipment selection, set-up and operation
111. indications, contraindications, hazards
112. application and/or modification
113. patient and equipment monitoring
114. various techniques used
115. cardiopulmonary resuscitation

Cardiorespiratory Monitoring:
116. ventilatory
117. electrocardiographic
118. hemodynamic

Cardiorespiratory Rehabilitation:
119. physical conditioning
120. home care planning
121. home care instruction
122. home care equipment and techniques
Clinical Assessment:
- 123. patient history and record
- 124. vital signs
- 125. inspection, palpation, percussion, and auscultation
- 126. determination of cardiorespiratory status

Continuous Ventilatory Support Techniques (Mechanical Ventilation, CPAP, PEEP):
- 127. equipment selection, set-up and operation
- 128. indications, contraindications, hazards
- 129. application and/or modification
- 130. patient and equipment monitoring
- 131. assessment for and techniques of weaning

Infection Control:
- 132. isolation and decontamination procedures
- 133. preparation for and techniques of sterilization
- 134. culturing techniques, significance of results

Laboratory Studies Assessment:
- 135. bacteriology
- 136. chemistries
- 137. hematology
- 138. radiology
- 139. special diagnostic procedures

Medical Gas Therapy (Oxygen, Helium, CO₂):
- 140. equipment selection, set-up and operation
- 141. indications, contraindications, hazards
- 142. application and/or modification
- 143. patient and equipment monitoring

Knowledge of Pathophysiology:
- 144. obstructive diseases, e.g., COPD, bronchitis, asthma
- 145. restrictive diseases, e.g., obesity, pulmonary fibrosis, thoracic cage deformity
- 146. CNS disorders, e.g., overdose, sleep apnea
- 147. pulmonary complications, e.g., atelectasis, pneumothorax
- 148. trauma, e.g., chest, head
- 149. acute parenchymal diseases, e.g., pneumonia, aspiration, ARDS
- 150. neuromuscular diseases, e.g., myasthenia gravis, Guillian Barré
- 151. cardiovascular disorders, e.g., cor pulmonale, left heart failure, pulmonary edema
- 152. pediatric/neonatal disorders, e.g., IRDS, cystic fibrosis, upper airway obstructions
Pharmacology:
- 153. indications, contraindications, side effects, and hazards
- 154. dosage and toxicity
- 155. routes of administration
- 156. monitoring patient response

Pulmonary Function Studies:
- 157. equipment selection, quality assurance
- 158. test selection and administration
- 159. blood gas sampling and analysis
- 160. test significance/interpretation of results

Others:
- 161. Medical Recordkeeping
- 162. Departmental/Administrative Recordkeeping
- 163. Physiology (Pulmonary, Cardiovascular, Renal, Acid-base, etc.)
- 164. Equipment Maintenance, Troubleshooting, and Repair
- 165. Administration/Supervision
- 166. Training/Education
APPENDIX E: Sample WEI Profile (Score Report)
### WORK EXPERIENCE INVENTORY

#### RESPIRATORY CARE PRACTITIONER

<table>
<thead>
<tr>
<th>JOB A</th>
<th>TITLE: CLINICAL SUPERVISOR RESPIRATORY CARE PRACTITIONER</th>
</tr>
</thead>
<tbody>
<tr>
<td>MONTHS EXPERIENCE:</td>
<td>36</td>
</tr>
<tr>
<td>LEVEL:</td>
<td>ADVANCED</td>
</tr>
<tr>
<td>POINTS:</td>
<td>546</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>JOB B</th>
<th>TITLE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MONTHS EXPERIENCE:</td>
<td></td>
</tr>
<tr>
<td>POINTS:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>JOB C</th>
<th>TITLE:</th>
</tr>
</thead>
<tbody>
<tr>
<td>MONTHS EXPERIENCE:</td>
<td></td>
</tr>
<tr>
<td>POINTS:</td>
<td></td>
</tr>
</tbody>
</table>

#### NAME: |

#### ADDRESS: |

#### PROCESSING DATE: 8/02/03 I.O.H.O. 92

#### CAREER INDEX: 5

#### AVERAGE RATING

<table>
<thead>
<tr>
<th>SITUATION</th>
<th>SIGNIFICANCE</th>
<th>IMPORTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (0-10)</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EXPOSURE TO WORK SITUATIONS</th>
<th>RATINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AVOIDING/INSTRUCTING OTHERS</td>
<td>7</td>
</tr>
<tr>
<td>2. INSTRUCTING CONTINUING ED/IN-SERVICE TRAINING</td>
<td>5</td>
</tr>
<tr>
<td>3. ATTENDING CONTINUING ED/IN-SERVICE TRAINING</td>
<td>4</td>
</tr>
<tr>
<td>4. INSTRUCTING PULMONARY/PEDIATRIC PERSONNEL RESOURCES</td>
<td>2</td>
</tr>
<tr>
<td>5. INSTRUCTING LEGAL/POLITICAL/PATIENT POLICIES</td>
<td>2</td>
</tr>
<tr>
<td>6. EVALUATING NEW PROCEDURES/TECHNOLOGIES</td>
<td>1</td>
</tr>
<tr>
<td>7. WORK IN ADULT INTENSIVE CARE</td>
<td>7</td>
</tr>
<tr>
<td>8. WORK IN PEDIATRIC/NEONATAL INTENSIVE CARE</td>
<td>0</td>
</tr>
<tr>
<td>9. WORK IN RECOVERY ROOM</td>
<td>2</td>
</tr>
<tr>
<td>10. WORK IN ADULT GENERAL CARE</td>
<td>3</td>
</tr>
<tr>
<td>11. WORK IN PEDIATRIC/NEONATAL GENERAL CARE</td>
<td>2</td>
</tr>
<tr>
<td>12. WORK IN EMERGENCY ROOM</td>
<td>4</td>
</tr>
<tr>
<td>13. WORK IN PULMONARY FUNCTION LAB</td>
<td>1</td>
</tr>
<tr>
<td>14. WORK IN OPERATING ROOM</td>
<td>1</td>
</tr>
<tr>
<td>15. WORK IN BLOOD GAS LAB</td>
<td>7</td>
</tr>
<tr>
<td>16. WORK IN RESEARCH LAB</td>
<td>1</td>
</tr>
<tr>
<td>17. WORK IN OTHER DIAGNOSTIC LAB</td>
<td>0</td>
</tr>
<tr>
<td>18. WORK IN OUT-PATIENT AREAS</td>
<td>3</td>
</tr>
<tr>
<td>19. WORK IN HOME CARE REHABILITATION</td>
<td>1</td>
</tr>
<tr>
<td>20. CLINICAL INSTRUCTION</td>
<td>5</td>
</tr>
<tr>
<td>21. EMERGENCY LAB-AIR TRANSPORT</td>
<td>3</td>
</tr>
<tr>
<td>22. EQUIPMENT MAINTENANCE AND REPAIR</td>
<td>4</td>
</tr>
<tr>
<td>POSITION RESPONSIBILITIES</td>
<td>AVERAGE RATING</td>
</tr>
<tr>
<td>-------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>1. Review existing data</td>
<td>3.3</td>
</tr>
<tr>
<td>2. Collect additional clinical data</td>
<td>3.6</td>
</tr>
<tr>
<td>3. Inspect chest x-ray</td>
<td>3.5</td>
</tr>
<tr>
<td>4. Perform bedside procedures</td>
<td>4.5</td>
</tr>
<tr>
<td>5. Recommend additional lab tests</td>
<td>3.7</td>
</tr>
<tr>
<td>6. Evaluate lab tests</td>
<td>3.6</td>
</tr>
<tr>
<td>7. Evaluate prescribed therapy and goals</td>
<td>3.0</td>
</tr>
<tr>
<td>8. Select and check equipment</td>
<td>5.6</td>
</tr>
<tr>
<td>9. Assure equipment cleanliness</td>
<td>2.6</td>
</tr>
<tr>
<td>10. Perform quality control</td>
<td>1.6</td>
</tr>
<tr>
<td>11. Explain planned therapy</td>
<td>1.5</td>
</tr>
<tr>
<td>12. Maintain patient airway</td>
<td>1.4</td>
</tr>
<tr>
<td>13. Remove bronchopulmonary secretions</td>
<td>1.3</td>
</tr>
<tr>
<td>14. Achieve adequate ventilation</td>
<td>1.3</td>
</tr>
<tr>
<td>15. Achieve adequate oxygenation</td>
<td>1.2</td>
</tr>
<tr>
<td>16. Evaluate patient response</td>
<td>1.2</td>
</tr>
<tr>
<td>17. Modify treatment</td>
<td>1.2</td>
</tr>
<tr>
<td>18. Initiate emergency resuscitation</td>
<td>1.2</td>
</tr>
<tr>
<td>19. Conduct rehabilitation and home care</td>
<td>1.2</td>
</tr>
<tr>
<td>20. Maintain records</td>
<td>1.2</td>
</tr>
<tr>
<td>21. Administer RT department</td>
<td>1.2</td>
</tr>
<tr>
<td>22. Supervise staff and services</td>
<td>1.2</td>
</tr>
<tr>
<td>23. Perform in-service training</td>
<td>1.2</td>
</tr>
<tr>
<td>24. Perform academic RT education</td>
<td>1.2</td>
</tr>
<tr>
<td>25. Assist physician</td>
<td>1.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESPIRATORY CARE KNOWLEDGE</th>
<th>AVERAGE RATING</th>
<th>NONE</th>
<th>MINOR</th>
<th>IMPORTANT</th>
<th>MOST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Airway care/maintenance</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Bronchial hygiene techniques</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Cardiopulmonary resuscitation</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Cardiorespiratory monitoring</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Cardiopulmonary rehabilitation</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Clinical assessment</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Continuous ventilatory support</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Infection control</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Laboratory studies assessment</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Medical gas therapy</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Pathophysiology</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Pharmacology</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Pulmonary function studies</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Medical recordkeeping</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Departmental recordkeeping</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Physiology</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Equipment maintenance</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Administration/supervision</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Training/educating</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F: Correlation Matrices for the Candidate and Study Data Sets
Correlation Matrix for the Candidate Data Set

<table>
<thead>
<tr>
<th></th>
<th>P1IGRAW</th>
<th>P10CRRAW</th>
<th>P21GRAW</th>
<th>P20DRRAW</th>
<th>P31GRAW</th>
<th>P302RAW</th>
<th>P41GRAW</th>
<th>P403RAW</th>
<th>P51GRAW</th>
<th>P504RAW</th>
<th>P61GRAW</th>
<th>P605RAW</th>
<th>P71GRAW</th>
<th>P706RAW</th>
<th>P81GRAW</th>
<th>P807RAW</th>
<th>P91GRAW</th>
<th>P908RAW</th>
<th>P10GRAW</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1IGRAW</td>
<td>1.0000</td>
<td>0.4133</td>
<td>0.3672</td>
<td>0.1538</td>
<td>2.6732</td>
<td>0.2148</td>
<td>0.1629</td>
<td>0.1277</td>
<td>0.2746</td>
<td>0.1245</td>
<td>1.0000</td>
<td>0.4133</td>
<td>0.3672</td>
<td>0.1538</td>
<td>2.6732</td>
<td>0.2148</td>
<td>0.1629</td>
<td>0.1277</td>
<td>0.2746</td>
</tr>
<tr>
<td>P10CRRAW</td>
<td>0.4133</td>
<td>1.0000</td>
<td>0.1723</td>
<td>0.1659</td>
<td>0.1284</td>
<td>0.1140</td>
<td>0.1977</td>
<td>0.1606</td>
<td>0.2561</td>
<td>0.1296</td>
<td>0.4133</td>
<td>1.0000</td>
<td>0.1723</td>
<td>0.1659</td>
<td>0.1284</td>
<td>0.1140</td>
<td>0.1977</td>
<td>0.1606</td>
<td>0.2561</td>
</tr>
<tr>
<td>P21GRAW</td>
<td>0.3672</td>
<td>0.1723</td>
<td>1.0000</td>
<td>0.2634</td>
<td>0.2364</td>
<td>0.1104</td>
<td>0.2774</td>
<td>0.2384</td>
<td>0.3438</td>
<td>0.2219</td>
<td>0.3672</td>
<td>0.1723</td>
<td>1.0000</td>
<td>0.2634</td>
<td>0.2364</td>
<td>0.1104</td>
<td>0.2774</td>
<td>0.2384</td>
<td>0.3438</td>
</tr>
<tr>
<td>P20DRRAW</td>
<td>0.1538</td>
<td>0.1659</td>
<td>0.2634</td>
<td>1.0000</td>
<td>0.2171</td>
<td>0.1271</td>
<td>0.2171</td>
<td>0.2171</td>
<td>0.3256</td>
<td>0.2923</td>
<td>0.1538</td>
<td>0.1659</td>
<td>0.2634</td>
<td>1.0000</td>
<td>0.2171</td>
<td>0.1271</td>
<td>0.2171</td>
<td>0.2171</td>
<td>0.3256</td>
</tr>
<tr>
<td>P31GRAW</td>
<td>2.6732</td>
<td>0.1284</td>
<td>0.2364</td>
<td>0.2171</td>
<td>1.0000</td>
<td>0.1629</td>
<td>0.1277</td>
<td>0.1277</td>
<td>0.2800</td>
<td>0.2630</td>
<td>2.6732</td>
<td>0.1284</td>
<td>0.2364</td>
<td>0.2171</td>
<td>1.0000</td>
<td>0.1629</td>
<td>0.1277</td>
<td>0.1277</td>
<td>0.2800</td>
</tr>
<tr>
<td>P302RAW</td>
<td>0.2148</td>
<td>0.1140</td>
<td>0.1104</td>
<td>0.1271</td>
<td>0.1629</td>
<td>1.0000</td>
<td>0.1977</td>
<td>0.1977</td>
<td>0.3116</td>
<td>0.2923</td>
<td>0.2148</td>
<td>0.1140</td>
<td>0.1104</td>
<td>0.1271</td>
<td>0.1629</td>
<td>1.0000</td>
<td>0.1977</td>
<td>0.1977</td>
<td>0.3116</td>
</tr>
<tr>
<td>P41GRAW</td>
<td>0.1629</td>
<td>0.1977</td>
<td>0.2774</td>
<td>0.2171</td>
<td>0.1629</td>
<td>0.1977</td>
<td>1.0000</td>
<td>0.3385</td>
<td>0.2034</td>
<td>0.2193</td>
<td>0.1629</td>
<td>0.1977</td>
<td>0.2774</td>
<td>0.2171</td>
<td>0.1629</td>
<td>0.1977</td>
<td>1.0000</td>
<td>0.3385</td>
<td>0.2034</td>
</tr>
<tr>
<td>P403RAW</td>
<td>0.1277</td>
<td>0.1277</td>
<td>0.2171</td>
<td>0.2171</td>
<td>0.1277</td>
<td>0.1277</td>
<td>0.3385</td>
<td>1.0000</td>
<td>0.3987</td>
<td>0.3634</td>
<td>0.1277</td>
<td>0.1277</td>
<td>0.2171</td>
<td>0.2171</td>
<td>0.1277</td>
<td>0.1277</td>
<td>0.3385</td>
<td>1.0000</td>
<td>0.3987</td>
</tr>
<tr>
<td>P51GRAW</td>
<td>0.2746</td>
<td>0.2800</td>
<td>0.2800</td>
<td>0.2800</td>
<td>0.2746</td>
<td>0.2800</td>
<td>0.3116</td>
<td>0.3987</td>
<td>1.0000</td>
<td>0.4097</td>
<td>0.2746</td>
<td>0.2800</td>
<td>0.2800</td>
<td>0.2800</td>
<td>0.2746</td>
<td>0.2800</td>
<td>0.3116</td>
<td>0.3987</td>
<td>1.0000</td>
</tr>
<tr>
<td>P504RAW</td>
<td>0.1245</td>
<td>0.2630</td>
<td>0.2630</td>
<td>0.2630</td>
<td>0.1245</td>
<td>0.2630</td>
<td>0.2923</td>
<td>0.3634</td>
<td>0.4097</td>
<td>1.0000</td>
<td>0.1245</td>
<td>0.2630</td>
<td>0.2630</td>
<td>0.2630</td>
<td>0.1245</td>
<td>0.2630</td>
<td>0.2923</td>
<td>0.3634</td>
<td>0.4097</td>
</tr>
<tr>
<td>P61GRAW</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>2.6732</td>
<td>1.0000</td>
<td>0.4133</td>
<td>0.3672</td>
<td>0.1538</td>
<td>2.6732</td>
<td>0.2148</td>
<td>0.1629</td>
<td>0.1277</td>
<td>0.2746</td>
</tr>
<tr>
<td>P605RAW</td>
<td>0.1245</td>
<td>0.2630</td>
<td>0.2630</td>
<td>0.2630</td>
<td>0.1245</td>
<td>0.2630</td>
<td>0.2923</td>
<td>0.3634</td>
<td>0.4097</td>
<td>0.1245</td>
<td>1.0000</td>
<td>0.4133</td>
<td>0.3672</td>
<td>0.1538</td>
<td>2.6732</td>
<td>0.2148</td>
<td>0.1629</td>
<td>0.1277</td>
<td>0.2746</td>
</tr>
</tbody>
</table>

Correlation Matrix for the Candidate Data Set
<table>
<thead>
<tr>
<th></th>
<th>P1ICERAN</th>
<th>P1DMKWA</th>
<th>P2ICERAN</th>
<th>P2DMKWA</th>
<th>P3ICERAN</th>
<th>P3DMKWA</th>
<th>P4ICERAN</th>
<th>P4DMKWA</th>
<th>P5ICERAN</th>
<th>P5DMKWA</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1ICERAN</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
<td>1.00000</td>
</tr>
<tr>
<td>P1DMKWA</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
<td>0.47200</td>
</tr>
<tr>
<td>P2ICERAN</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
</tr>
<tr>
<td>P2DMKWA</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
<td>0.24762</td>
</tr>
<tr>
<td>P3ICERAN</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
</tr>
<tr>
<td>P3DMKWA</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
<td>0.91847</td>
</tr>
<tr>
<td>P4ICERAN</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
</tr>
<tr>
<td>P4DMKWA</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
<td>0.30316</td>
</tr>
<tr>
<td>P5ICERAN</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
</tr>
<tr>
<td>P5DMKWA</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
<td>0.58107</td>
</tr>
</tbody>
</table>

Correlation Matrix for the Study Data Set