

CLINICAL INSTRUCTORS' INTERRATER RELIABILITY  
ON SELECTED PHYSICAL THERAPY SKILLS

---

A THESIS  
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF MASTER OF SCIENCE  
IN THE GRADUATE SCHOOL OF THE  
TEXAS WOMAN'S UNIVERSITY

INSTITUTE OF HEALTH SCIENCES  
SCHOOL OF HEALTH CARE SERVICES  
HEALTH SCIENCES INSTRUCTION PROGRAM

BY  
JUDITH ANN HEMBREE, B.S.

---

DENTON, TEXAS

AUGUST 1984

The Graduate School  
Texas Woman's University  
Denton, Texas

July 20 1984

We hereby recommend that the thesis prepared under

our supervision by Judith Ann Hembree

entitled "Clinical Instructors' Interrater Reliability on  
Selected Physical Therapy Skills"

be accepted as fulfilling this part of the requirements for the Degree of Master of Science

Committee:

Barbara J. Chamer  
Chairman  
Robert Fuley  
Eileen Morrison

Accepted:

Leslie M. Thompson  
Provost of the Graduate School

## TABLE OF CONTENTS

	Page
LIST OF TABLES . . . . .	vi
Chapter	
I. INTRODUCTION . . . . .	1
Statement of the Problem . . . . .	1
Statement of Purposes. . . . .	2
Hypotheses . . . . .	2
Definition of Terms. . . . .	3
Assumptions. . . . .	5
Limitations. . . . .	6
Significance of the Study. . . . .	7
Summary. . . . .	8
II. REVIEW OF LITERATURE . . . . .	9
The Process of Evaluation. . . . .	10
Clinical Evaluation Instruments. . . . .	12
The Mastery and Assessment of Clinical Skills . . . . .	13
The Use of Videotape in Evaluation . . . . .	14
Statistical Reliability. . . . .	16
Summary. . . . .	17
III. METHODOLOGY. . . . .	18
Setting. . . . .	18
Population and Sample. . . . .	19
Protection of Human Subjects . . . . .	19
Instruments. . . . .	20
Mastery and Assessment of Clinical Skills. . . . .	20
Development of the Videotape . . . . .	20
Pilot Study. . . . .	22
Findings from the Experts. . . . .	23
Data Collection. . . . .	24
Treatment of Data. . . . .	24

Chapter	Page
IV. FINDINGS . . . . .	26
Description of Participants. . . . .	26
Analysis of the Data . . . . .	26
Hypothesis 1 . . . . .	27
Hypothesis 2 . . . . .	30
Hypothesis 3 . . . . .	33
Hypothesis 4 . . . . .	33
Results from the Item Analysis . . . . .	34
Summary. . . . .	37
V. SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS. . . . .	38
Summary. . . . .	38
Conclusions. . . . .	39
Discussion . . . . .	40
Statistical Limitations. . . . .	40
Videotape Limitations. . . . .	41
Division of Opinion among the Experts. . . . .	42
Experience Level of the Clinical Instructor . . . . .	42
Item Analysis of the <u>MACS</u> . . . . .	43
Limitations of the <u>MACS</u> . . . . .	43
Recommendations. . . . .	45
APPENDIXES	
A. Skill 10: Muscle Testing. . . . .	46
B. Skill 11: Goniometry. . . . .	48
C. Consent Form . . . . .	50
D. Oral Directions to the Clinical Instructors On the Videotape . . . . .	52
E. Letters of Permission to Conduct the Study from the Physical Therapy Department Directors. . . . .	57
F. Pilot Study: Expert Responses by Skills 10 and 11 Key Indicators . . . . .	60
G. Videotape Evaluation: Responses by the Experts. . . . .	63
H. Pilot Study: Expert Comments and Responses by Skills 10 and 11 Key Indicators . . . . .	65
I. Muscle Test and Range of Motion Forms from the Videotape . . . . .	70
J. Instructions for Rating the Student in the MACS . . . . .	75



	Page
REFERENCE NOTE . . . . .	79
REFERENCES . . . . .	80
BIBLIOGRAPHY . . . . .	84

## LIST OF TABLES

Table	Page
1. Responses of Full-Time Clinical Instructors by Skill 10 Key Indicators . . . . .	28
2. Responses of Part-Time Clinical Instructors by Skill 10 Key Indicators . . . . .	29
3. Responses of Full-Time Clinical Instructors by Skill 11 Key Indicators . . . . .	31
4. Responses of Part-Time Clinical Instructors by Skill 11 Key Indicators . . . . .	32
5. Item Analysis Results for Skill 10 . . . . .	35
6. Item Analysis Results for Skill 11 . . . . .	36

## CHAPTER I

### INTRODUCTION

Clinical evaluation is an integral component in the education of physical therapy students. The Mastery and Assessment of Clinical Skills (MACS) evaluation instrument was developed to improve the consistency and quality of clinical evaluation in Texas. A large number of academic and clinical physical therapists assisted in the development of the MACS in order to insure content validity.

This study was undertaken to determine the interrater reliability of a group of physical therapy clinical instructors who have supervised full-time and part-time physical therapy students on their clinical affiliations. Two psychomotor skills were presented on videotape. After viewing each skill, the clinical instructors then rated the student's performance in the MACS. Interrater reliability of the clinical instructors was then determined.

#### Statement of the Problem

The interrater reliability of the physical therapy clinical instructors utilizing the MACS was unknown. The interrater reliability scores of selected physical therapy psychomotor skills were analyzed. This study attempted to

answer the question: What are the Baylor Health Care System physical therapy clinical instructors' interrater reliability scores on selected psychomotor skills?

### Statement of Purposes

The primary purpose of the study was to determine the Baylor Health Care System physical therapy clinical instructors' interrater reliability scores on the two psychomotor skills from the Mastery and Assessment of Clinical Skills evaluation instrument. The second purpose was to make a videotape of student performance on two selected psychomotor skills from the MACS instrument. The third purpose was to review and evaluate the written comments on each skill to determine the reasons why the ratings were given.

### Hypotheses

For the purposes of this study, the four null hypotheses were:

1. There is no significant interrater reliability in the ratings obtained from the Baylor Health Care System physical therapy clinical instructors on each key indicator of skill 10 (muscle testing) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

2. There is no significant interrater reliability in the ratings obtained from the Baylor Health Care System physical therapy clinical instructors on each key indicator of skill 11 (goniometry) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

3. There is no significant difference in the interrater reliability score between full-time and part-time physical therapy clinical instructors in the Baylor Health Care System on each key indicator of skill 10 (muscle testing) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

4. There is no significant difference in the interrater reliability score between full-time and part-time physical therapy clinical instructors in the Baylor Health Care System on each key indicator of skill 11 (goniometry) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

#### Definition of Terms

For the purposes of this study, the following operational definitions were used.

Clinical instructor. The licensed physical therapist who has primary responsibility for evaluating a physical therapy student in the clinic.

Full-time clinical instructor. A clinical instructor who supervises only full-time, 5 to 8 week affiliation students.

Part-time clinical instructor. A clinical instructor who supervises only part-time, 1 or 2 week affiliation students.

Mastery and Assessment of Clinical Skills (MACS).  
The evaluation instrument developed by the Texas Consortium for Physical Therapy Clinical Education that is used to assess Texas physical therapy students during their clinical affiliations.

Skill 10 (muscle testing). One of the 54 skills in the third edition of the MACS. Each skill addresses an area of competence in physical therapy practice. This skill describes the patient evaluation procedure "manual muscle testing." See Appendix A for a full statement of the skill.

Skill 11 (goniometry). This skill describes the patient evaluation procedure "range of motion" testing

using a goniometer. See Appendix B for a full statement of the skill.

Key indicator. One of the performance criteria for a skill in the MACS.

Baylor Health Care System. A group of hospitals in North Texas, with Baylor University Medical Center in Dallas serving as the central administrative unit.

#### Assumptions

Eight assumptions were made for this study.

1. The MACS consists of a group of skills which represent the minimum competencies necessary for entry level physical therapy practice.

2. Content validity of the MACS has been established through an ongoing review process by physical therapists throughout the state of Texas.

3. The physical therapy clinical instructors in the Baylor Health Care System have received adequate training in the use of the MACS to use it effectively in the clinic and in this study.

4. Videotaping presented the psychomotor skills in a realistic manner.

5. Each clinical instructor had an equal opportunity to rate performance while viewing the videotape of each skill.

6. Each physical therapy clinical instructor was able to see the videotape clearly.

7. Clinical instructors rated performance on a videotape in a similar manner to performance in the clinic.

8. The change in physical environment did not bias the experts or participants.

#### Limitations

The results of this study were evaluated within the context of the following limitations.

1. The extent of experience in use of the MACS was variable among the physical therapy clinical instructors.

2. The extent of training in use of the MACS was variable among the physical therapy clinical instructors.

3. The size of the convenient sample was small, and only two hospitals within the Baylor Health Care System were utilized. This limits the generalizability of the study.

4. There was no way to control personal conversations about the videotapes during the study.



### Significance of the Study

The objectives of the Baylor Health Care System include serving as an educational center for health care personnel and serving as a research center. This study has contributed to the completion of both of these objectives, and it has provided a preliminary examination of the effectiveness of two skills in the MACS, a physical therapy clinical evaluation instrument which is used by physical therapists at Baylor.

Clinical education is a major activity of the physical therapy departments in the Baylor Health Care System. A large percentage of the physical therapists participate in student supervision as full-time or part-time clinical instructors. It is hoped that this study may have contributed to identification of the need for more extensive training in the use of the MACS for physical therapy clinical instructors.

The MACS has been the clinical evaluation instrument for Texas physical therapy students since 1979. There have been only limited attempts to study interrater reliability on a few skills (Dragotta, Note 1). It is, therefore, difficult to assess student competence in the

clinic when clinical instructors may be interpreting the MACS differently.

Physical therapists must have a reliable means of determining student clinical competence. The MACS was designed for this purpose. This study explored the interrater reliability of two psychomotor skills in a hospital setting.

#### Summary

The problem of determining interrater reliability was studied in the physical therapy clinical education program of the Baylor Health Care System. The evaluation instrument utilized was the MACS. Interrater reliability was assessed on two psychomotor skills, muscle testing (10) and goniometry (11).

## CHAPTER II

### REVIEW OF LITERATURE

In this chapter, four topics relevant to this study are discussed: the process of evaluation, clinical evaluation instruments, the use of videotape in evaluation, and statistical reliability. Numerous studies have been conducted in the fields of psychology and education on evaluation and statistical analysis of reliability. In several fields, videotape has been utilized as the medium for presenting the experimental subjects for study. A review of past physical therapy journals revealed many attempts to create an evaluation instrument applicable to all physical therapy students. This literature review provided a representative sample of the pertinent studies in each area.

#### The Process of Evaluation

Moore and Perry (1976) defined evaluation as "The appraisal of the worth of a person, place, or thing in terms of internal or external criteria" (p. 3). They also stated that evaluation may be performed to predict performance; describe a program, person, or thing; or prescribe further actions.

Evaluation may be completed by a supervisor (or teacher), a peer, or by the person being evaluated (self-evaluation). Friesen and Dunning (1973) found that psychology students could effectively rank order the performances of their peers presented on a videotape. Their ratings were significantly higher than those of their supervisors. They concluded that peer evaluation was a useful adjunct to other forms of evaluation.

Several studies have been completed on the problem of errors in evaluations utilizing rating scales. Littlefield, Anthracite, Herbert, and McKendree (1983), in a clinical study involving medical students, proposed that a handicap score be given to faculty raters, based on their tendency to make these errors. The student's final grade or assessment would then be based on this adjusted rating. Moore and Perry (1976), in a non-clinical study, discussed errors of leniency and central tendency, the halo effect, logical error, proximity error, and contrast error. An error of leniency occurred when the rater gave the student too high a rating. An error of central tendency was the tendency to use primarily the middle of a rating scale. The halo effect involved giving a higher rating to a student who has done well in other areas. A logical error occurred when similar traits are rated the

same. A contrast error involved the tendency to rate a student's behavior at opposite ends of the continuum of behavior. A proximity error involved similar ratings of nearby items.

Gibb (1983) studied the halo effect further. He found that grades on term papers correlated significantly with course grades only when information about the student's previous academic performance was available to the grader. He suggested that to increase objectivity, term papers should be graded without knowledge of the student's past performance.

Marston, Zimmerer, and Vaughn (1978) described the problem of coder drift in the evaluation of teacher performance. They defined coder drift as, "Change that takes place in the way an observation system is used after coders have been trained" (p. 1). They stressed the need for retraining to normative standards in order to increase reliability of ratings.

In summary, evaluation is a complex process that has many purposes, takes many forms, and may contain several types of errors. In clinical evaluation, an appropriate instrument must be developed to complete the process effectively.

### Clinical Evaluation Instruments

Many health professions have developed evaluation instruments and rating scales to assess student performance. Erviti, Fabrey, and Bunce (1979) developed a simple rating scale of 46 items for faculty to use in assessing medical students. The responses called for were seldom, sometimes, usually, and consistently. The middle responses tended to be used the most.

The American Occupation Therapy Association (AOTA) developed an evaluation form for students completing their clinical affiliations. The Field Work Performance Report has been adopted by AOTA as the official instrument for evaluation of clinical performance. Crocker, Muthard, Slaymaker, and Samson (1975) discussed the process of establishing validity and reliability of the instrument. Overall validity was determined by relating the ratings on the Field Work Performance Report to a rating based on potential for being hired in a facility. The assumption made was that a student who performed well would have good potential for employment. Clinicians and educators assisted in item writing and reviewing the material for content validity. Interrater reliability was assessed by having two staff members rate every student on the pilot study.

Kerr and Mickelson (1971) described a checklist and rating scale form. The form consisted of general statements about clinical performance and a list of common physical therapy procedures. Dickinson, DiMarino, and Pfitzenmaier (1973) described a descriptive evaluation form developed by several schools in New York. The objectives for this form must be developed by each clinical center. Wilhelm (1969) proposed a Q-methodology for developing a clinical evaluation instrument.

#### The Mastery and Assessment of Clinical Skills

The MACS evaluation instrument was developed in 1977 through 1979 by the Texas Consortium for Physical Therapy Clinical Education. The MACS consists of 54 skills, with performance criteria (key indicators) for each skill. The MACS is utilized by physical therapy students in all the schools in Texas, as well as some schools in other states.

A large number of clinical and academic physical therapists representing many practice areas in Texas participated in writing the MACS. Thus, content validity may be reasonably assumed. Initially, each skill was reviewed by three groups of physical therapists before it

was placed in the MACS. In addition, the MACS has undergone three revisions.

Dragotta, Academic Coordinator of Clinical Education at the University of Texas Health Science Center at San Antonio, has performed some studies on reliability. She indicated that consistency of training is a major factor in consistency of ratings (Dragotta, Note 1).

#### The Use of Videotape in Evaluation

Videotape has been utilized as a medium for presenting the experimental subjects in several studies. Nugent and Labs (1978) suggested that videotaping was the one method by which behavior can be held constant. This allowed determination of the consistency of raters' judgments. Their study involved performance of complex electrical work by Navy personnel.

Person, Klein, Hyman, and Cook (1977) performed an elaborate study of transcultural interrater reliability using videotape. Psychiatrists in Turkey and Missouri viewed schizophrenic patients on videotape and noted presence or absence of symptoms. This study allowed psychiatrists from two distant countries to observe and rate patients from their own and another culture. This



was logistically possible only through the use of videotape.

Piercy and Laird (1983) utilized videotaped segments of activity to develop a short rating scale for effectiveness of family therapists. Curlee (1981) utilized videotape recordings to study reliability of ratings of disfluency and stuttering. Rafael and Marinoff (1973) advocated use of videotape by teachers to reobserve children in a classroom.

Boykin and Nelson (1981) studied the effects of instructions and calculation procedures on observers' accuracy, agreement, and calculation correctness. Pairs of students watched videotapes of children. It was found that instructions that emphasized the importance of obtaining high interobserver agreement yielded higher agreement scores than accuracy scores. In addition, when students were allowed to perform their own calculations, they erroneously inflated agreement levels for their own data and deflated agreement levels for data contrived to appear collected by other students. The authors concluded that directions should emphasize accuracy and not agreement.

In summary, videotape has been used in a wide range of studies. It is a convenient method that may be applied to a variety of questions. Directions given for a videotape study appeared to affect the results.

### Statistical Reliability

Studies in the fields of education (Burton, 1981), psychology (Naglieri & Maxwell, 1981), speech therapy (Norris, Harden, & Bell, 1980), medical education (Markert & Shoes, 1980), physical education (Godbout & Schutz, 1983), and many others have considered the problem of determining reliability. There are many statistical studies dealing with a variety of tests. Determination of reliability is affected by the statistical test selected.

Overall and Klett (1972) described interrater reliability in terms of ratings made independently by two observers. The two-rater format was frequently used in interrater reliability studies. Bartko (1976) concluded that a high intraclass correlation reliability coefficient is obtained when there is small within-subjects variance. Bremman and Prediger (1981) discussed using kappa as a descriptive statistic summarizing agreement between two judges across a number of objects. Green (1981) compared

three measures of agreement: proportion of agreement, G-index, and kappa. He concluded that no single agreement index is appropriate for all sets of data. Smith (1979) developed a computer program for determining agreement of paired raters.

Littlefield, Murrey, and Garman (1977) described use of a generalizability coefficient which corrects ratings for the number of raters and trials. Bintig (1980) and Burton (1981) advised against the use of the statistical test kappa for multiple raters. However, Bartko and Carpenter (1976) distinguished different forms of kappa, and related generalized kappa as an appropriate test for multiple raters.

### Summary

The need for a reliable clinical evaluation instrument in physical therapy and in other professions has been documented. Many statistical tools have been devised to test reliability. Studies have shown videotape to be a useful medium for experiments on reliability. This study attempted to use this information to test interrater reliability of clinical instructors in a hospital setting.

## CHAPTER III

### METHODOLOGY

This study utilized a quasi-experimental, one-group posttest only design. There was no control group.

#### Setting

The physical therapy clinical instructors viewed the videotapes in the physical therapy gymnasium at Baylor Institute for Rehabilitation (formerly Swiss Avenue Hospital) and in two conference rooms at Baylor Hospital. The videotapes were shown during regularly scheduled in-service meetings. The chairs and monitor were placed in a similar pattern each time. It was not possible to reserve the same room for each presentation.

Make-up sessions were held in the office of the researcher. One person attended a make-up session for each videotape.

Three experts (two clinical and one academic) viewed the videotapes in the researcher's office. The two clinical experts viewed the videotapes at the same time. One academic expert viewed the videotapes in a classroom at the Texas Woman's University Presbyterian campus with only the researcher present.

### Population and Sample

All 18 licensed physical therapy clinical instructors in the Baylor Health Care System participated in this study. This group formed a convenient sample of the population of physical therapy clinical instructors in Texas. Of this sample, 7 physical therapy clinical instructors worked with full-time students and 11 worked with part-time students. The physical therapy clinical instructors' experience in utilizing the MACS was unknown.

### Protection of Human Subjects

The patient and student "actors" in the videotapes were volunteers. A thorough description of the study was presented and a rehearsal of the videotapes completed before the subjects signed a release form (Appendix C).

Participation in the study by the clinical instructors was voluntary. Instructions on the videotape informed the clinical instructors that turning in the written materials constituted participation consent (Appendix D).

Permission was obtained from the Directors of the Physical Therapy Departments to conduct the study. Copies of these letters are in Appendix E.

### Instruments

#### Mastery and Assessment of Clinical Skills

The evaluation instrument utilized was the third edition of the MACS. This instrument was required for use by Texas physical therapy students. It was considered to be part of the public domain and was available for use by clinical instructors. It was not necessary to secure permission to use this instrument.

Skills 10 (muscle testing) and 11 (goniometry) were selected for this study for several reasons. They were basic physical therapy skills which all students must master. Both skills required close visual inspection by the clinical instructor rating a student; this made them ideal for videotaping. Skill 10 (muscle testing) also had a strong tactile component. The investigator was interested in how the clinical instructors would react to not being able to touch the patient. Finally, there was a patient volunteer available who demonstrated obvious problems related to these two skills.

#### Development of the Videotape

Using skills 10 and 11 as a guideline, a plan was developed for each videotape using a 5 x 8 card system.

Each card had three elements: visual elements and camera angle, topic to be presented, and script for the student and patient. The script included both correct and incorrect responses by the student, based on the performance criteria (key indicators) in the MACS.

The student "actor" reviewed muscle testing and goniometry with the investigator. The student rehearsed twice with the investigator and once with the patient "actor."

The videotape was filmed in the physical therapy gymnasium at Baylor Institute for Rehabilitation. Filming was done on Sunday when the physical therapy gymnasium was not in use. The camera person was a staff physical therapist at Baylor Institute for Rehabilitation who had developed the departmental videocassette library. Equipment used included a JVC color videocamera (model KY-1900U), a JVC recorder (model CR-6060U), and 3/4-inch FUJI videocassette tape.

During the actual filming, there was not exact adherence to the script because of the anxiety level of the student and some unplanned responses by the patient. For example, the patient developed a cramp in the hamstring muscles during the knee flexion measurement on

skill 11. The student had to spontaneously modify the procedure to accommodate for this problem. The main ideas stated in the script were portrayed in the final videotape. A copy of the videotape was placed in the library of the Texas Woman's University.

### Pilot Study

Before the licensed physical therapy clinical instructors viewed the videotapes, they were viewed by four experts. Two of the experts were from an academic setting, and two were from a physical therapy clinical setting. Their experience level as physical therapists ranged from 8 to 24 years. All four experts had several years experience with the MACS, and two participated in the development of it.

The experts viewed the videotapes and rated the student in the same manner as the clinical instructors. Responses by the experts to skills 10 and 11 are given in Appendix F. The experts also completed an evaluation form on the acceptability of the videotapes (Appendix G). The experts were interviewed by the researcher as to the reason for their ratings that indicated less than acceptable performance, and these responses are found in Appendix H.



The responses by the experts formed the basis for determining the correct answer on each key indicator. The researcher made the final decision on each rating, but at least 50% of the experts were in agreement with the decision.

#### Findings from the Experts

The experts were divided in their opinion of the correct answer on approximately 50% of the key indicators. There was not a pattern of response in terms of the background of the expert, academic or clinical. There was greater division of opinion on skill 11 than on skill 10.

An example of disagreement among the experts was found on skill 10, key indicator "d," "Instructs patient in motion to be done--demonstrates, if necessary." Two experts (one academic and one clinical) rated the student's performance acceptable (yes). One clinical expert stated performance was inconsistent and assigned a rating of question mark (?). One academic expert stated the student had little affect and the patient looked bored and rated the student's performance not acceptable (no).

No changes were made in the videotape or procedures based on the findings from the experts. All four experts

agreed that the videotapes were acceptable for use in the study.

### Data Collection

The licensed physical therapy clinical instructors viewed the videotapes during two regular in-service meetings (two at Baylor Institute for Rehabilitation and two at Baylor Hospital). They were given a copy of each skill and paper for notes at the beginning of each session. They were given a copy of the student's written work (produced on the videotape) at the end of the videotape (Appendix I).

The physical therapy clinical instructors rated the student on each skill using the standard directions (Appendix J) for recording in the MACS. The physical therapy clinical instructors turned in their skill ratings form and notes at the end of the meeting.

### Treatment of Data

Tabulation of the frequency of yes, no, question mark, and blank responses on each key indicator of each skill was done. The written comments on each skill were reviewed and evaluated by the researcher to determine why "no" ratings were given. These comments are discussed in the following chapter.

Reliability was determined using the statistical test, generalized Kappa (Bartko & Carpenter, 1976). This test was appropriate for an experiment utilizing multiple raters with multiple questions.

Item analysis was performed on each key indicator of skills 10 and 11. A t-test was performed to determine if there was a significant difference between interrater reliability scores of physical therapy clinical instructors working with full-time vs. part-time students.

## CHAPTER IV

### FINDINGS

The results of the study are described in this chapter. Tables of data are presented as well as a narrative description.

#### Description of Participants

There were 18 licensed physical therapy clinical instructors in the study. Seven of the physical therapy clinical instructors worked with full-time students, and 11 of the clinical instructors worked with part-time students. Of the seven full-time clinical instructors, three were from Baylor Institute for Rehabilitation and four were from Baylor Hospital; both are part of the Baylor Health Care System. Of the 11 part-time clinical instructors, 5 were from Baylor Institute for Rehabilitation and 6 were from Baylor Hospital.

#### Analysis of the Data

The data were described for each hypothesis and for the item analysis. Several tables of data are provided.

### Hypothesis 1

Hypothesis 1 stated: There is no significant inter-rater reliability in the ratings obtained from the Baylor Health Care System physical therapy clinical instructors on each key indicator of skill 10 (muscle testing) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument. The data for skill 10 are presented in Tables 1 and 2. Analysis of the data utilizing the statistical test, generalized Kappa, revealed a reliability score of .337 for the full-time physical therapy clinical instructors and .355 for the part-time physical therapy clinical instructors. Hypothesis 1 was accepted.

Generalized Kappa ranges from  $-.143$  to  $1$ , with  $.43$  as the midpoint. Thus, the reliability scores obtained on this skill were very low. This indicated that the results were due to chance rather than to agreement among the physical therapy clinical instructors.

Table 1

Responses of Full-Time Clinical Instructors  
by Skill 10 Key Indicators

Key Indicator	Response			Question Mark (?)
	Yes (✓)	No (N)	Blank (No answer)	
a	7			
b	3	2		2
c		5	2	
d	7			
e		5	1	1
f	6			1
g	7			
h	1	4	2	
i	7			
j	4	2	1	
k	3	3		1

N = 7.

Table 2

Responses of Part-Time Clinical Instructors  
by Skill 10 Key Indicators

Key Indicator	Response			Question Mark (?)
	Yes (✓)	No (N)	Blank (No answer)	
a	11			
b	3	5		3
c	1	8	1	1
d	11			
e	6			5
f	9	2		
g	11			
h	3	4		4
i	11			
j	3	6		2
k	6	2		3

N = 11.

## Hypothesis 2

Hypothesis stated: There is no significant inter-rater reliability in the ratings obtained from the Baylor Health Care System physical therapy clinical instructors on each key indicator of skill 11 (goniometry) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument. The data for skill 11 are presented in Tables 3 and 4. Analysis of the data utilizing the statistical test, generalized Kappa, revealed a reliability score of .230 for the full-time physical therapy clinical instructors and .220 for the part-time physical therapy clinical instructors. Hypothesis 2 was accepted.

The reliability scores obtained on this skill were also very low. This indicated that the results were due to chance rather than to agreement among the physical therapy clinical instructors.



Table 3

Responses of Full-Time Clinical Instructors  
by Skill 11 Key Indicators

Key Indicator	Response			Question Mark (?)
	Yes (✓)	No (N)	Blank (No answer)	
a	7			
b	6	1		
c	2	5		
d	7			
e	3	3		1
f	6			1
g	4			3
h	2	2		3
i		3	3	1
j	4	2		1

N = 7.

Table 4  
Responses of Part-Time Clinical Instructors  
by Skill 11 Key Indicators

Key Indicator	Response			Question Mark (?)
	Yes (✓)	No (N)	Blank (No answer)	
a	7	2		2
b	6	4		1
c	1	5		5
d	10	1		
e	7	2		2
f	8	2		1
g	9	2		
h	5	6		
i	2	4	4	1
j	9	2		

N = 11.

### Hypothesis 3

Hypothesis 3 stated: There is no significant difference in the interrater reliability score between full-time and part-time physical therapy clinical instructors in the Baylor Health Care System on each key indicator of skill 10 (muscle testing) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument. A  $t$ -test was performed to test this hypothesis. The results showed  $t = 1.151$ . This hypothesis was accepted.

In order for  $t$  to be significant at the .05 level,  $t_{16}$  must equal 2.12 (18 subjects from 2 groups with 2 degrees of freedom). Therefore, there is no significant difference in the reliability coefficient between the two groups of physical therapy clinical instructors.

### Hypothesis 4

Hypothesis 4 stated: There is no significant difference in the interrater reliability score between full-time and part-time physical therapy clinical instructors in the Baylor Health Care System on each key indicator of skill 11 (goniometry) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument. A  $t$ -test was performed to test this hypothesis. The results showed  $t = .667$ . Therefore, there is

no significant difference in the reliability coefficient between the two groups of physical therapy clinical instructors. This hypothesis was accepted.

### Results from the Item Analysis

An item analysis was performed to determine the percentage of correct responses on each key indicator of skills 10 and 11. The percentages ranged from 27.8 to 100.0 for skill 10 and from 11.1 to 94.4 for skill 11. Tables 5 and 6 present the percentages of correct responses for each of the key indicators as well as the index of discrimination and the index of difficulty for each key indicator.

For skill 10, the index of discrimination ranged from -0.200 to 0.800; and for skill 11, the index of discrimination ranged from -0.200 to 0.600. A positive index of discrimination indicated a well-written key indicator, one which would be judged correctly by those clinical instructors with the highest percentage of correct answers. For example, key indicator "c" on skill 10 stated, "Takes part passively through range to check range of motion." This key indicator was clearly written and received an index of discrimination score of 0.800 (Table 5).

Table 5  
Item Analysis Results for Skill 10

Key Indicator	% Correct Responses	Index of Difficulty	Index of Discrimination
a	100.0	0.000	0.000
b	33.3	0.900	-0.200
c	72.2	0.400	0.800
d	100.0	0.000	0.000
e	61.1	0.500	1.000
f	83.3	0.100	-0.200
g	100.0	0.000	0.000
h	44.4	0.500	1.000
i	100.0	0.000	0.000
j	44.4	0.500	0.600
k	27.8	0.500	0.600

Table 6  
Item Analysis Results for Skill 11

Key Indicator	% Correct Responses	Index of Difficulty	Index of Discrimination
a	11.1	0.900	0.200
b	66.7	0.400	-0.400
c	55.6	0.500	0.200
d	94.4	0.000	0.000
e	55.6	0.400	0.400
f	11.1	0.900	-0.200
g	16.7	0.800	0.400
h	44.4	0.700	-0.200
i	38.9	0.600	0.400
j	72.2	0.300	0.600

For both skills 10 and 11, the index of difficulty ranged from 0.000 to 0.900. An index of difficulty score of 0.500 indicated a well-written key indicator. For example, key indicator "c" on skill 11 stated, "Exposes the joint and its landmarks." This key indicator received an index of difficulty score of 0.500. This indicated differentiation between the highest scoring 27% of the group of clinical instructors and the lowest scoring 27% of the clinical instructors (Table 6).

#### Summary

The results of the study supported all four hypotheses. There was no significant interrater reliability among the full-time or part-time physical therapy clinical instructors on skills 10 or 11. There was no significant difference in reliability between the two groups.

## CHAPTER V

### SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

#### Summary

This study determined the Baylor Health Care System physical therapy clinical instructor interrater reliability scores on two psychomotor skills from the Mastery and Assessment of Clinical Skills evaluation instrument. A videotape was developed for each skill. Eighteen physical therapy clinical instructors then rated the student actor on the videotape utilizing the MACS evaluation instrument. The results were tallied and statistical tests were performed to accept or reject the hypotheses. All four null hypotheses listed below were accepted.

1. There is no significant interrater reliability in the ratings obtained from the Baylor Health Care System physical therapy clinical instructors on each key indicator of skill 10 (muscle testing) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

2. There is no significant interrater reliability in the ratings obtained from the Baylor Health Care System physical therapy clinical instructors on each key



indicator of skill 11 (goniometry) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

3. There is no significant difference in the interrater reliability score between full-time and part-time physical therapy clinical instructors in the Baylor Health Care System on each key indicator of skill 10 (muscle testing) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

4. There is no significant difference in the interrater reliability score between full-time and part-time physical therapy clinical instructors in the Baylor Health Care System on each key indicator of skill 11 (goniometry) as measured by the third edition of the Mastery and Assessment of Clinical Skills evaluation instrument.

### Conclusions

The results showed poor interrater reliability among the physical therapy clinical instructors in the Baylor Health Care System for skill 10 (muscle testing) and skill 11 (goniometry). Interrater reliability was slightly less among the full-time clinical instructors than part-time

clinical instructors. However, there was no statistical difference between the two groups on both skills.

### Discussion

There were many possible reasons why interrater reliability was not at an acceptable level on the skills studied. These are discussed below.

### Statistical Limitations

There were 18 participants in this study, divided into 2 groups of 7 full-time and 11 part-time physical therapy clinical instructors. There were 11 key indicators on skill 10 (muscle testing) and 10 key indicators on skill 11 (goniometry). Thus, each participant responded to 21 questions (key indicators). For each skill, there were more key indicators than there were participants in each group of physical therapy clinical instructors. This made statistical analysis of the results difficult: there was a large number of questions and a small number of participants to answer them.

In addition, only .04% of the 54 affective and psychomotor skills in the MACS were studied. The results of this study cannot be generalized to the other skills in the MACS.

### Videotape Limitations

Utilizing videotape as the method to present the skills may have been an influence on the results of the study. The skills were illustrated; however, the filming was not all of high quality. The photographer had some experience but was not an expert. Some of the close-up shots of the student's written work were poorly focused. However, a copy of this work was given to the clinical instructors at the end of the videotape. It was not necessary to read the written work on the videotape.

The physical therapy clinical instructors were unable to "touch" the patient. Some clinical instructors may believe that the tactile sense or "hands on" approach is necessary to test the patient to determine if the student was accurate. When the clinical instructor is unsure of the patient's response, he or she will often perform the procedure on the patient to check the accuracy of the student's work.

The physical therapy clinical instructors were unable to interact with the student actor on the videotape. In actual clinical situations, the clinical instructor usually discusses the procedure with the student as well as observing it. Knowing the student's thought process

may change the clinical instructor's assessment of his or her performance.

Overall, videotaping may not be the ideal method to present the skills. Alternate methods of measuring interrater reliability may be more accurate.

#### Division of Opinion among the Experts

The experts, academic and clinical, were unable to agree on the correct rating for the student on many of the key indicators. This made assessment of interrater reliability difficult because it was questionable what the "correct" rating should be on many of the key indicators.

In this study, individual accuracy by the physical therapy clinical instructors was necessary in order to achieve interrater reliability for the entire group. If the physical therapy clinical instructors did not select the "correct" answer, then interrater reliability was low.

#### Experience Level of the Clinical Instructor

There was wide variation among the clinical instructors' experience in utilizing the MACS. It appears that a clinical instructor's concept of the MACS may

evolve over time; this may also decrease interrater reliability. Continued in-service education on use of the MACS may help solve this problem if differences in perception of these skills really exist.

#### Item Analysis of the MACS

The item analysis provided an additional way of looking at the MACS. In an ideal study when interrater reliability is high, the percentage of correct responses would be 100%. This would cause the index of difficulty and the index of discrimination to be zero.

#### Limitations of the MACS

A review of the key indicators for skills 10 (muscle testing) and 11 (goniometry) reveals that many of them have more than one concept. For example, key indicator "b" of skill 10 states, "Positions the part to determine the degree of strength; i.e., anti-gravity or gravity assisted position and eliminates substitutions when they occur." There are at least two thoughts within one statement. A review of the comments written by the experts and the participants on each skill revealed different ratings based on response to different aspects of the key indicator.

When the key indicators expressed only one concept, there was much greater agreement among participants. For example, key indicator "g" on skill 10 states, "Checks both the normal and involved side." There was 100% agreement on this key indicator.

This evaluation discrepancy among clinical instructors on some key indicators suggests that the MACS may need further revision for clarity. The results from the item analysis support this need for revision.

The low interrater reliability scores in this study indicate the need for continued investigation of the process of evaluating physical therapy students in the clinic. If clinical instructors are unable to agree on a student's performance, then the faculty would not know when entry level competencies were met; thus, graduation may need to be postponed until all clinical faculty were in agreement. In order to ensure adequate clinical education for the student and the possession of adequate clinical skills for safe practice, physical therapists must be able to reliably assess student competence in the clinic.

### Recommendations

There are several possible further studies that may be done to determine interrater reliability. The number of skills may be increased in a study utilizing the same format. Additional demographic information may be included to help isolate the factors that decrease interrater reliability. A selected number of skills could be studied using a different methodology.

The evaluation instrument may need to be changed. Clarification of the criteria for assigning a rating in the MACS may be studied and appropriate changes made.

In-service education may be needed to improve utilization of the MACS. The videotapes produced for this study could be utilized to teach physical therapy clinical instructors how to rate the student in the MACS.

## APPENDIXES



## APPENDIX A

### SKILL 10: MUSCLE TESTING

## SKILL # 10

Student's Name \_\_\_\_\_

Performs accurate measurements of muscle strength in patients with varied neuromuscular and musculoskeletal disorders.

Affiliation Number								Performance Criteria: Key Indicators of Acceptable Mastery
1	2	3	4	5	6	7	8	
S								a. Selects the appropriate type of test for the specific order.
CI								
S								b. Positions the part to determine the degree of strength; ie. antigravity or gravity assisted position and eliminates substitutions when they occur.
CI								
S								c. Takes part passively through range to check range of motion.
CI								
S								d. Instructs patient in motion to be done--demonstrates, if necessary.
CI								
S								e. Palpates tendons and muscle bellies for evidence of contraction.
CI								
S								f. Gives resistance correctly (location and amount).
CI								
S								g. Checks both the normal and involved side.
CI								
S								h. Judges patient's muscle strength accurately.
CI								
S								i. Performs the evaluation within a reasonable length of time.
CI								
S								j. Modifies test procedure appropriately without sacrificing accuracy when pain, limitation of motion, or psychological reactions are limiting factors.
CI								
S								k. Records the results of the muscle test accurately and concisely noting any deviations from the accepted testing procedures.
CI								

Does this student do other things which you feel indicate performance beyond the minimum needed for safe and effective practice or fail to do things which you feel indicate deficiencies in this area? If so, please note affiliation number, describe, and initial.

## APPENDIX B

### SKILL 11: GONIOMETRY

## SKILL # 11

Student's Name \_\_\_\_\_

Evaluates joint range of motion by goniometric measurement.

Affiliation Number								Performance Criteria: Key Indicators of Acceptable Mastery
1	2	3	4	5	6	7	8	
S								a. Explains the procedure to the patient.
CI								
S								b. Positions patient in preferred or acceptable alternative position.
CI								
S								c. Exposes the joint and its landmarks.
CI								
S								d. Instructs the patient to perform the motion actively, if possible.
CI								
S								e. Maintains alignment of the goniometer with landmarks and axis of motion.
CI								
S								f. Reads the scale at the beginning and end of the range.
CI								
S								g. Eliminates substitutions.
CI								
S								h. Judges patient's range of motion accurately.
CI								
S								i. Differentiates between limitation due to joint restriction or soft tissue disorders.
CI								
S								j. Records results and notes if motion was active or passive.
CI								

Does this student do other things which you feel indicate performance beyond the minimum needed for safe and effective practice or fail to do things which you feel indicate deficiencies in this area? If so, please note affiliation number, describe, and initial.

APPENDIX C

CONSENT FORM

## TEXAS WOMAN'S UNIVERSITY

## CONSENT FORM

We, the undersigned, do hereby consent to the recording of our voices and/or images by Judith Ann Hembree, acting on this date under the authority of the Texas Woman's University. We understand that the material recorded today may be made available for educational, informational, and/or research purposes; and we do hereby consent to such use.

We hereby release the Texas Woman's University and the undersigned parties acting under the authority of the Texas Woman's University from any and all claims arising out of such taping, recording, reproducing, publishing, transmitting, or exhibiting as is authorized by the Texas Woman's University.

## SIGNATURES OF PARTICIPANTS\*

_____	_____ Date
_____	_____ Date
_____	_____ Date

\*\*\*

The above consent form was read, discussed, and signed in my presence. In my opinion, the person(s) signing said consent form did so freely and with full knowledge and understanding of its contents.

_____	_____ Date
Witness	

Note: This form was signed by the student actor and patient in the videotape and was filed with the Human Subjects Review Committee.

APPENDIX D

ORAL DIRECTIONS TO THE CLINICAL INSTRUCTORS  
ON THE VIDEOTAPE

ORAL DIRECTIONS TO THE CLINICAL INSTRUCTORS  
ON THE VIDEOTAPE

I wish to welcome you to this study. I appreciate your donation of your time and your expertise. I hope you will enjoy this opportunity to participate in clinical educational research.

The "Blue MACS," developed by the Texas Consortium for Physical Therapy Clinical Education, has been used in Texas and in the Baylor Health Care System since 1978. The purpose of this study is to evaluate the physical therapy clinical instructor's utilization of the "Blue MACS" for assessment of student skills. The study results will also be used to complete the requirements for the degree of Master of Science in Health Sciences Instruction at the Texas Woman's University.

Participation in this study is voluntary. You will not be compensated in any way for participation or for any injury that may be a result of participation in this study. Turning in the written material at the end of this meeting will constitute consent for participation.

You will not be identified in any way in this study. If you write comments on the pages of the "Blue MACS"



provided to you, please use only the initials CI at the end of the comment.

Now, let's discuss the background for the videotape patient. Mr. Jones is a 42-year-old male who has been referred to physical therapy for reassessment of lower extremity weakness and range of motion deficits. He has been assigned to your student.

In 1973, Mr. Jones jumped off a roof and fractured his right lower extremity. There were many vascular complications, and an open reduction internal fixation was performed twice. Mr. Jones received 4 months of physical therapy as an outpatient in 1974.

Your student has performed other range of motion and muscle tests and has asked to be checked off on these two skills when he assesses Mr. Jones. The student has completed self-evaluation in the "Blue MACS" and has placed a check by all key indicators for these two skills.

As you watch your student on the videotape, you may want to write some notes on the student's performance on the blank page provided to you.

At the end of the videotape you will be provided a copy of the assessment results as recorded by the student.

You may review this in the same way that you would review any work your student places in the chart.

After you have viewed the videotape and reviewed the written work, please record your assessment of the student's performance on the page of the "Blue MACS" provided to you. Follow the standard instructions for recording in the "Blue MACS."

Before you turn in the page from the "Blue MACS," please check full-time or part-time clinical instructor in the box at the bottom of the page from the "Blue MACS." You are a full-time clinical instructor if your last student was full-time or a 5- or 6-week affiliation student. You are a part-time clinical instructor if your last student was a preclinical or 1- or 2-week affiliation student.

The patient assessment by your student is about to begin. If you are unable to see this monitor clearly, please reposition yourself now. Thank you.

The treatment portion of the videotape follows. The last three paragraphs of directions follow the test on the videotape. This paragraph will not be on the videotape.

Your student has now completed the patient assessment, and it is time for you to evaluate. Remember to

follow the standard instructions for recording in the "Blue MACS." That is, for each key indicator, record a check mark for yes, N for no, question mark for uncertain, or a blank for no opportunity to observe. Remember to check the box for full-time or part-time clinical instructor at the bottom of the page of the "Blue MACS."

Please do not discuss this videotape with anyone in this room or with any therapist during the day for the next week. You will be given an opportunity to discuss the study when it is complete. Please do not discuss it until that time.

Thank you for your participation. Please record your assessment of the student now.

APPENDIX E

LETTERS OF PERMISSION TO CONDUCT THE STUDY FROM THE  
PHYSICAL THERAPY DEPARTMENT DIRECTORS



**BAYLOR UNIVERSITY MEDICAL CENTER**

Boone Powell, Jr., F.A.C.H.A.  
President

Carl P. Collins Hospital  
Karl and Esther Hoffmann Memorial Hospital  
Erik and Margarete Jonsson Medical and Surgical Hospital  
George W. Truett Memorial Hospital  
Mintie S. Neal Teaching and Research Hospital

**April 2, 1984**

**Ms. Judy Hembree**  
Physical Therapy Department  
3504 Swiss Avenue  
Dallas, Texas 75204

Dear Ms. Hembree:

You have my permission to conduct your study concerning physical therapy clinical instructors interrater reliability during two of our regularly scheduled Monday in-service meetings. As you know, our staff is interested in clinical education and I am sure would actively participate as subjects. I understand, however, that participation is voluntary.

You are scheduled to conduct the study for the staff at Jonsson and Collins on April 16 and 23 1984. The meeting lasts from 8:15 to 9:00 A.M.

Sincerely,

A handwritten signature in cursive script that reads "Kay Youens".

Kay Youens  
Director, Physical Therapy

KEY/der



## SWISS AVENUE HOSPITAL

3504 Swiss Avenue  
Dallas, Texas 75204  
(214) 826-7030

M. Tim Parris  
Executive Director

## Rehabilitation

W. L. Parker, M.D.  
Medical Director

R. L. Bruce, M.D.  
Associate Director

April 2, 1984

Ms. Judy Hembree  
Physical Therapy Department

Dear Judy:

You have my permission to conduct your study concerning physical therapy clinical instructors' interrater reliability during two of our regularly scheduled in-service meetings. As you know, the staff is very interested in clinical education and I am sure would actively participate as subjects. I understand, however, that participation is voluntary.

You are scheduled to conduct the study on April 9 and 12, 1984. The meetings are scheduled from 8:15 to 9:00 a.m.

Sincerely,

*LaNelle Riester, PT*

LaNelle Riester, P.T.  
Director, Physical Therapy

## APPENDIX F

PILOT STUDY: EXPERT RESPONSES BY SKILLS 10 AND 11  
KEY INDICATORS

## PILOT STUDY: EXPERT RESPONSES BY SKILL 10

## KEY INDICATORS

Key Indicator	Response			Question Mark (?)
	Yes (✓)	No (N)	Blank (No answer)	
a	C,D,B			A
b	C,D	B,A		
c		C,D,B,A		
d	C,A	D		B
e		C,D,B,A		
f	C,B	D		A
g	C,D,B,A			
h		D,B,A	C	
i	C,D,B,A			
j		C,D,A	B	
k	D	C,A		B

Note: Key  
 A, B, C, D refer to the individual experts  
 (N=4).  
 A and B are clinical experts.  
 C and D are academic experts.



## PILOT STUDY: EXPERT RESPONSES BY SKILL 11

## KEY INDICATORS

Key Indicator	Response			Question Mark (?)
	Yes (✓)	No (N)	Blank (No answer)	
a	A	C,D		B
b	B,D	C		A
c	B	A,C,D		
d	A,B,C,D			
e	C,D	A		B
f	C,D			A,B
g	C	D		A,B
h		A,B,C,D		
i		A,D	B	C
j	A,C,D			B

Note: Key  
 A, B, C, D refer to the individual experts  
 (N=4).  
 A and B are clinical experts.  
 C and D are academic experts.

## APPENDIX G

VIDEOTAPE EVALUATION: RESPONSES BY THE EXPERTS

# VIDEOTAPE EVALUATION

Please rate the components of the videotape listed below as: acceptable for use in the study, not acceptable for use in the study, or undecided. The comments section may be used for suggested changes or any other ideas you wish to suggest.

Component	Not		
	Acceptable	Acceptable	Uncertain
Purpose of the study. Comments:	A,B,C,D		
Background information on the patient. Comments: D--Why does he need to be reassessed?	A,B,C		D
Explanation of written materials. Comments: B--Unclear about the ROM form-- can't see it; maybe say that it will be given to CI	A,C,D		B
Directions for the study. Comments:	A,B,C,D		
Directions for noting full- or part-time clinical instructor. Comments: C--Might want to know # of years experience individual has using MACS	A,B,D		C
Other (specify) Comments: A--ROM form--distracting to show-- made me worry about not being able to see it.	A		

## APPENDIX H

PILOT STUDY: EXPERT COMMENTS AND RESPONSES BY  
SKILLS 10 AND 11 KEY INDICATORS

## PILOT STUDY: EXPERT COMMENTS AND RESPONSES BY SKILL 10 KEY INDICATORS

Key Indicator	Response	Expert	Comments
a	?	A	Should do a sitting test for the feet. Trying a functional (standing) test for plantarflexion was inappropriate because of knee weakness; should try in parallel bars.
b	N	A	Did not do; try another position
	N	B	Did not eliminate substitutions. Did not stabilize most of time.
c	N	A	Never did.
	N	B	Never did.
	N	C	Checked range of motion in previous video, however, did not passively take through range in this instance.
	N	D	No comment.
d	?	B	Inconsistent instructions.
	N	D	Little affect; patient looked bored.
e	N	A	Not consistent.
	N	B	Not consistent.
	N	C	"Touched" the right ankle, however, did not really palpate tendons or muscle belly.
	B	D	No comment.
f	?	A	Cannot tell on video.
	N	D	Used long lever at hip--if do Kendall, use a spring test. Abduction--with a painful knee should not cross joint. Gastroc-soleus--should be more thorough.

Key Indicator	Response	Expert	Comments
h	N	A	One action only is a gross test.
	N	B	One action only is a gross test.
	x	C	Must touch patient to decide. Possibly was trace on foot.
	N	D	On the foot, did not check the muscles properly. Do not cross the joint.
j	N	A	When there was pain or no motion, gave up.
	x	B	Could not tell.
	N	C	Did not alter position for knee flexion when cramp developed.
	N	D	Did not try to find out about quad. Did not pursue raising up on toe.
k	N	A	Gross test; did motion once. Some may have been right.
	?	B	Could be no. Recorded for gross test.
	N	C	Should have palpated tibialis anterior better and placed in gravity eliminated position.

N = 4.

Note. N = No; x = Blank; ? = Question Mark; All other responses were yes.

## PILOT STUDY: EXPERT COMMENTS AND RESPONSES BY SKILL 11 KEY INDICATORS

Key Indicator	Response	Expert	Comments
a	?	B	Did patient already know why he was doing the test?
	N	C	Be more specific in explanation (i.e., as to purpose of measuring joint range of motion, reason for measuring both sides, etc.).
	N	D	Overall, little explanation to the patient. Patient looked puzzled and bored. Affective skills no.
b	?	A	Could be no. Want to see again. Thomas test--film from side of plinth so if not limited, could see. Never asked about back comfort, especially in prone.
	N	C	Should position patient differently (supine or sitting) to measure knee flexion after hamstring cramp developed on left. Also try to support lower extremity in measuring hip abduction/adduction.
c	N	A	Did not uncover trochanter.
	N	C	Exposed landmarks for ankle/knee but not hip joint and trunk.
	N	D	No comment.
e	N	A	Possibly ? External rotation sitting--goniometer rode up.
	?	B	Greater trochanter. External rotation sitting left leg--let ride up.
f	?	A	Unsure--would turn goniometer. Not recorded on form.
	?	B	Put down only end range--appeared to do it.

Key Indicator	Response	Expert	Comments
g	?	A	Could be no. Tried, but did not use hands to stabilize. Did not do enough, but did some.
	?	B	Inconsistent--did some verbally.
	N	D	Did not eliminate external rotation with abduction. Thomas test--did not lock pelvis.
h	N	A	Did every motion once--do 2 or 3. ? + 1 on knee extension.
	N	B	No starting range. Some active and some passive--did not compare.
	N	C	Record ROM to nearest 5°. Read the goniometer off the part in abduction/adduction measurement.
	N	D	See g. Left knee flexion done too quickly.
i	N	A	Did not record ankle
	x	B	No comment.
	?	C	Could be blank.
	N	D	Not recorded or mentioned.
j	?	B	Inconsistent--do passive on more and see what find.

N = 4.

Note. N = No; x = Blank; ? = Question Mark; All other responses were yes.



APPENDIX I

MUSCLE TEST AND RANGE OF MOTION FORMS  
FROM THE VIDEOTAPE

BAYLOR UNIVERSITY MEDICAL CENTER  
RANGE OF MOTION - LOWER EXTREMITY

Name Videos Jones Age 42 Diagnosis S/P (R) L/E Fracture

Positioning of Patient for Testing of Joints:

Supine: Hip Flexion (opposite hip extended), Abduction, Adduction; Knee Extension;  
Dorsiflexion, Plantarflexion, Inversion, Eversion.

Prone: Hip Extension; Knee Flexion.

Sitting: Hip Internal & External Rotation; Knee Extension.

Please Note Below If Patient Positioned Otherwise.

Also Note Positioning For Knee Extension.

A = Active

P = Passive

LEFT					RIGHT				
				VS	Examiner's Initials	VS			
				4-1	Date	4-1			
					HIP				
A				108	Flexion	0-120	120		A
P									P
A				-5	H. Extension	0-15	-5	THOMAS TEST	A
P									P
A				45	Abduction	0-45	22		A
P									P
A				12	Adduction	0-30	21		A
P									P
A				20	Int. Rot.	0-45	22		A
P									P
A				21	Ext. Rot.	0-45	20		A
P									P

## KNEE

A			115	Flexion	0-135	72				A
P						64				P
A			+1	Extension	0	+3				A
P										P
				ANKLE						
A			15	Dorsiflexion	0-20					A
P						-15				P
A			45	Plantarflexion	0-50	40				A
P										P
A			45	Inversion	0-30					A
P						20				P
A			55	Eversion	0-20					A
P						15				P

Comments:

BAYLOR UNIVERSITY MEDICAL CENTER  
MANUAL MUSCLE TEST

Name Vides Jones Age 42 Diagnosis s/p R L/E Fracture

N Normal Completes range of motion against gravity plus maximal manual resistance  
 G+ Good plus Completes range of motion against gravity plus moderate to maximal resistance  
 G Good Completes range of motion against gravity plus moderate manual resistance  
 G- Good minus Completes range of motion against gravity plus minimal to moderate resistance  
 F+ Fair plus Completes range of motion against gravity plus any minimal resistance  
 F Fair Completes range of motion against gravity  
 F- Fair minus Completes more than  $\frac{1}{2}$  range of motion against gravity  
 P+ Poor plus Initiation of range of motion against gravity to  $\frac{1}{2}$  the range  
 P Poor Completes range of motion gravity eliminated  
 P- Poor minus Any partial range of motion gravity eliminated  
 T Trace Contraction with no joint motion  
 O Zero No evidence of contraction

Grades Fair (F) and below in red

Motion				Muscle	
Left	Right			Left	Right
V5	V5			V5	V5
4-1	4-1	Examiner		4-1	4-1
		Date			
		NECK	FLEXION	Accessory (C2-3) Sternocleidomastoid	
			EXTENSION	Accessory (C1-6) Extensor Group	
		TRUNK	FLEXION	(T8-12) Rectus Abdominis	
			ROTATION	(T8-10) External Oblique	
				(T10-L1) Internal Oblique	
			EXTENSION	(T1-9) Thoracic Extensors	
				(T10-L5) Lumbar Extensors	
		PELV. ELEV.		(T12-L2) Quadratus Lumborum	



APPENDIX J

INSTRUCTIONS FOR RATING THE STUDENT IN THE MACS

## V. ASSESSING STUDENT PERFORMANCE

In recording your assessment of students' mastery of clinical skills, please remember that for ALL STUDENTS, regardless of their level of training and/or experience (ie. part-time or full-time affiliations), the STANDARD FOR APPROVAL IS THE SAME:

THE LEVEL OF PERFORMANCE NECESSARY FOR SAFE AND EFFECTIVE PRACTICE AS A GRADUATE PHYSICAL THERAPIST.

Please do not let your natural hope that many students achieve excellence distract you from the PURPOSE OF THE MACS:

TO ASSESS THE STUDENT'S PROGRESS TOWARD ADEQUATE MASTERY OF ESSENTIAL SKILLS

### A. RECORDING ON THE INDIVIDUAL SKILL SHEETS

1. Assess ONLY those competencies that you try to help the student master during this affiliation. LEAVE THE OTHER PAGES BLANK.
2. Assess each KEY INDICATOR SEPARATELY. Each describes a different criterion that is part of the entire skill, and the student may do well at some and require additional practice with others.
3. RECORD your assessment of the key indicator (criteria) ON THE INDIVIDUAL SKILL SHEETS in the STUDENT's COPY of the Blue MACS. This should be done in the numbered column designated for your affiliation, using the following:

## RATING SCALE

- (✓) = STUDENT HAS MET MINIMUM LEVEL CRITERIA  
 Place a check mark opposite a key indicator as soon as you are satisfied with the student's performance. Your "Yes" (✓) rating may be recorded whenever you feel the student is able to perform this aspect of a skill independently at the minimal level necessary for safe and effective practice in your setting—as early as the first day in some cases. This step should NOT be left until mid or final evaluation sessions or the Formative Evaluation opportunities will be lost.
- (N) = STUDENT HAS NOT YET MET ENTRY LEVEL CRITERIA  
 Write (N) opposite the key indicator if the student is not doing this well enough or consistently enough at this time for you to feel his/her work would be safe and effective if it were unsupervised. In other words, the student still needs to improve his/her performance of the particular key indicator.
- (?) = Record a (?) ONLY IF YOU HAVE NOT HAD SUFFICIENT OPPORTUNITY TO OBSERVE THE STUDENT'S PERFORMANCE OF THE KEY INDICATOR IN ORDER TO MAKE A DECISION.
- (Leave a BLANK) = IF YOU HAVE NOT BEEN WORKING WITH THE STUDENT ON MASTERY OF THE KEY INDICATOR.  
 This is equivalent to: "no opportunity to perform" OR "not applicable" in your setting.

Do NOT feel that because an instructor at an earlier affiliation assessed the student's performance as acceptable, you cannot give an "N" rating. A student's ability to perform may be determined in part by special characteristics of the setting in which he/she is working. Your rating should be based on how the student does with the type of patients, facilities, procedures and staff as your setting provides.

#### 4. USE OF COMMENT SECTION of the skill sheet:

Use the spaces at the foot of each skill sheet page to comment on any aspects of the student's performance you feel are important.

INDICATE AFFILIATION NUMBER AND PLACE YOUR INITIALS BESIDE EACH COMMENT YOU MAKE.



SITUATIONS DESERVING COMMENTS include: (see following examples)

1. ANY "N's" YOU RECORD REQUIRE A COMMENT which details why the performance is not at minimum entry level. (See Affil. #1 comment below)
2. Qualify the conditions under which you recorded a "Yes" rating. (See Affil. #2 comment below)
3. Comment on performance which exceeds a minimum achievement. (See Affil. #3 comment below)

REMEMBER, the "Yes" rating simply says that the student is at least adequate. He/she may or may not perform at a higher than minimum level and deserves that feedback.

EXAMPLES: Skill #11 Goniometry									
Affiliation Number									
	1	2	3	4	5	6	7	8	Performance Criteria: Key Indicators of Acceptable Mastery
S									1. Differentiates between limitation due to joint restriction or soft tissue disorders.
CI	N	✓	✓						
<p>Does this student do other things which you feel indicate performance beyond the minimum for safe and effective practice or fail to do things which you feel indicate deficiencies in this area? If so, please <u>note affiliation number, describe, and date.</u></p> <p>Affil #1 8/22/81 MK Inconsistent in differentiating, OK at times</p> <p>Affil #2 11/5/81 CF Performed well but only on hip and knee limitations.</p> <p>Affil #3 2/28/82 DP Shows high degree of discrimination and plans corrective program accordingly.</p>									

Note. From Mastery and Assessment of Clinical Skills (3rd ed.) by Texas Consortium for Physical Therapy Clinical Education, 1982.

#### REFERENCE NOTE

1. Dragotta, N. Personal communication, March 27, 1983.

## REFERENCES

- Bartko, J. J. On various intraclass correlation reliability coefficients. Psychological Bulletin, 1976, 83, 762-765.
- Bartko, J. J., & Carpenter, W. T. On the methods and theory of reliability. The Journal of Nervous and Mental Disease, 1976, 163, 307-317.
- Bintig, A. The efficiency of various estimations of reliability of rating scales. Educational and Psychological Measurement, 1980, 40, 619-643.
- Boykin, R. A., & Nelson, R. O. The effects of instructions and calculation procedures on observers' accuracy, agreement, and calculation correctness. Journal of Applied Behavior Analysis, 1981, 14, 479-489.
- Brennan, R. L., & Prediger, D. J. Coefficient Kappa: Some uses, misuses, and alternatives. Educational and Psychological Measurement, 1981, 41, 687-699.
- Burton, N. Estimating scorer agreement for nominal categorization systems. Educational and Psychological Measurement, 1981, 41, 953-962.
- Crocker, L. M., Muthard, J. E., Slaymaker, J. E., & Samson, L. A performance rating scale for evaluating clinical competence of occupational therapy students. The American Journal of Occupational Therapy, 1975, 29(2), 81-86.
- Curlee, R. F. Observer agreement of disfluency and stuttering. Journal of Speech and Hearing Research, 1981, 24, 595-600.
- Dickinson, R., DiMarino, J., & Pfitzenmaier, J. A common evaluation instrument. Physical Therapy, 1973, 53, 1075-1080.

- Erviti, V., Fabrey, L. J., & Bunce, J. V. The development of rating scales to assess the performance of medical students in clinical settings. Philadelphia: National Board of Medical Examiners, 1979. (ERIC Document Reproduction Service No. ED 186 458)
- Friesen, D. D., & Dunning, G. B. Peer evaluation and practicum supervision. Counselor Education and Supervision, 1973, 12, 229-235.
- Gibb, G. D. Influence of "halo" and "demon" effects in subjective grading. Perceptual and Motor Skills, 1983, 56(1), 670-70.
- Godbout, P., & Schutz, R. W. Generalizability of ratings of motor performances with reference to various observational designs. Research Quarterly for Exercise & Sport, 1983, 54(1), 20-27.
- Green, S. B. A comparison of three indexes of agreement between observers: Proportion of agreement, G-index, and Kappa. Educational and Psychological Measurement, 1981, 41, 1069-1072.
- Kern, B. P., & Mickelson, J. M. The development and use of an evaluation instrument for clinical education. Physical Therapy, 1971, 51, 540-546.
- Littlefield, J. H., Anthracite, N. E., Herbert, R., & McKendree, J. Adjusting observational ratings to improve inter-rater consistency. San Antonio, TX: University of Texas Health Science Center at San Antonio, 1983. ERIC Document Reproduction Service No. ED 229 420)
- Littlefield, J. H., Murrey, A. J., & Garman, R. E. Assessing the generalizability of clinical rating scales. San Antonio, TX.: University of Texas Health Science Center at San Antonio, 1977. (ERIC Document Reproduction Service No. ED 139 830)
- Markert, R. J., & Shores, J. H. Assuring fairness in the medical school admission interview through analysis of rater difficulty and consistency. Fort Worth, TX.: Office of Medical Education, Texas College of Osteopathic Medicine, 1980. (ERIC Document Reproduction Service No. ED 185 114)

- Marston, P. T., Zimmerer, L. K., & Vaughn, W. C. Coder drift: A reliability problem for teacher observations. Austin, TX.: Research and Development Center for Teacher Education, 1978. (ERIC Document Reproduction Service No. ED 161 897 95)
- Mays, M. J. Reliability of a method of evaluation the clinical performance of a physical therapy student. Physical Therapy, 1973, 53, 1298-1306.
- Moore, M. L., & Perry, J. F. Clinical education in physical therapy: Present status/future needs. Washington, D.C.: Section for Education, American Physical Therapy Association, 1976.
- Naglieri, J. A., & Maxwell, S. Inter-rater reliability and concurrent validity of the Goodenough-Harris and McCarthy draw-a-child scoring systems. Perceptual and Motor Skills, 1981, 53, 343-348.
- Norris, M., Harden, J., & Bell, D. M. Listener agreement on articulation errors of four- and five-year-old children. Journal of Speech and Hearing Disorders, 1980, 45, 378-389.
- Nugent, W. A., & Laabs, G. J. Performance test objectivity: Comparison of interrater reliabilities of three observation formats. San Diego, CA.: Navy Personnel Research and Development Center, 1978. (ERIC Document Reproduction Service No. ED 186 474)
- Overall, J. E., & Klett, J. C. Applied multivariate analysis. New York: McGraw-Hill Book Co., 1972.
- Person, T., Klein, H. E., Hyman, A., & Cook, N. T. Transcultural interrater reliability. Journal of Social Service Research, 1977, 1, 187-203.
- Piercy, F. P., & Laird, R. A. A family therapist rating scale. Journal of Marital and Family Therapy, 1983, 9(1), 4-59.
- Rafael, B., & Marinoff, S. L. Using videotape for teacher training. Young Children, 1973, 28(4), 217-219.

Smith, D. U. Estimating the reliability of clinical ratings: An alternative approach. Detroit, MI.: Division of Educational Services and Research, Wayne State University School of Medicine, 1979. (ERIC Document Reproduction Service No. ED 174 680)

Texas Consortium for Physical Therapy Clinical Education. Mastery and Assessment of Clinical Skills (3rd ed.). Galveston, TX.: University of Texas Medical Branch Student Bookstore, 1981.

Wilhelm, I. J. Q-Methodology in rating the clinical competency of physical therapy students: A model. Physical Therapy, 1969, 49, 1227-1230.

## BIBLIOGRAPHY

American Academy of Orthopaedic Surgeons. Joint motion: Method of measuring and recording. Chicago: Author, 1965.

Daniels, L., Williams, M., & Worthingham, C. Muscle testing: Techniques of manual examination. Philadelphia: W. B. Saunders, 1956.

Kendall, H. O., Kendall, F. P., & Wadsworth, G. E. Muscles: Testing and function (2nd ed.). Baltimore: Williams & Wilkins, 1971.