Test-Retest Reliability of the Levels of Rehabilitation Scale IIB

Maura Katherine Murphy
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Rehabilitation and Therapy Commons

Recommended Citation
Murphy, Maura Katherine, "Test-Retest Reliability of the Levels of Rehabilitation Scale IIB" (1991). Master's Theses. 979.
https://scholarworks.wmich.edu/masters_theses/979
TEST-RETEST RELIABILITY OF THE LEVELS OF REHABILITATION SCALE IIB

by

Maura Katherine Murphy

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the requirements for the Degree of Master of Science
Department of Occupational Therapy

Western Michigan University
Kalamazoo, Michigan
April 1991
The purpose of this study was to evaluate test-retest reliability of the Levels of Rehabilitation Scale IIB (LORS IIB), developed by Carey and Posavac in 1982. Reliability is an important concept when a therapist begins to assess a client's activities of daily living skills. If a test is reliable, the therapist will have a dependable means by which to measure client progress with stability and accuracy.

This study was conducted with eight volunteer post-stroke victims. The study found that the LORS IIB assessment was highly significant in test-retest reliability.
ACKNOWLEDGEMENTS

I would like to express my appreciation to Barbara Hemphill, Cindee Peterson, and Dr. Molly Vass for their help in the preparation of this study.

Maura Katherine Murphy
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Test-retest reliability of the levels of rehabilitation scale II

Murphy, Maura Katherine, M.S.
Western Michigan University, 1991
NOTE TO USERS

THE ORIGINAL DOCUMENT RECEIVED BY U.M.I. CONTAINED PAGES WITH POOR PRINT. PAGES WERE FILMED AS RECEIVED.

THIS REPRODUCTION IS THE BEST AVAILABLE COPY.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .......................................................... ii

CHAPTER

I. INTRODUCTION .................................................................. 1
   Rationale ........................................................................ 1
   Assumptions About Reliability ........................................ 2
   Reliability in Activity of Daily Living Assessments ............ 3
   Statement of the Problem ................................................ 4
   Summary ........................................................................ 4

II. LITERATURE REVIEW ......................................................... 6
   Internal Consistency ....................................................... 6
   Alternate Forms ............................................................. 6
   Inter-rater Reliability ..................................................... 7
   Test-retest Reliability .................................................... 7
   Test-retest Examination ................................................ 7
   Summary ....................................................................... 9

III. METHODOLOGY ............................................................... 11
   Subjects .......................................................................... 11
   Instrument ................................................................. 11
   Procedure .................................................................... 12
   Research Question ......................................................... 12
   Data Analysis ............................................................... 13
Table of Contents—Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV. DISCUSSION</td>
<td>15</td>
</tr>
<tr>
<td>V. CONCLUSION</td>
<td>17</td>
</tr>
</tbody>
</table>

APPENDICES

A. Letter of Consent from Human Subjects Institutional Review Board. 18
B. Permission To Use LORS IIB Assessment 20
C. Letter of Informed Consent 22

BIBLIOGRAPHY 24
CHAPTER I

INTRODUCTION

The purpose of this study was to evaluate test-retest reliability of the Levels of Rehabilitation Scale IIB (LORS IIB) developed by Carey and Posovac in 1982. Reliability is an important concept when a therapist begins to assess a client for activities of daily living skills. If a test is reliable, the therapist has a dependable means to measure client progress with stability and accuracy.

Rationale

In the literature on reliability, two main theories have been proposed about statistical error. The first is the generalization theory which postulates that there is a universe of observations, or conditions of the measurement which may influence the results of a test. The conditions the tester chooses to be accepted then define the ability to generalize the datum to the universe. Reliability, under this theory, is the accuracy of the generalization or generalizability of the datum to every situation or condition of the test (Cronbach, Gleser, Nanda, Rajaratnam, 1972).

Reliability, in this study, is based on the traditional error theory. This theory states that there is no interaction between the subject and the test; rather, differences come from true subject variations combined with random variations among observations (error).
or the components of the situation being measured. Although some conditions of the testing situation are controlled, reliability is based on the observed score with its inherent error component.

Assumptions About Reliability

Inherent in the theory are several assumptions about the test score. Assumption one is that scores are generally fallible. Each test score is the summation of the true component (score under ideal conditions) and the error component (Kerlinger, 1986). Reliability, therefore, depends on the amount of error component in the test score.

Another assumption is that the error can be either positive or negative and that in the larger population the mean is zero. In other words, within scores one subject may obtain a score of 12 and an error component of -1, while another obtains a score of 11 and an error component of 2. Over the entire population, however, the mean of the error component is zero.

One last assumption is that in a large population errors are uncorrelated with true scores. In other words, high true scores do not have either more positive or more negative error scores (Guilford, 1954). For example, if one subject has a score of 12 and an error component of 1, this does not necessarily mean that another subject with a score of 3 has a lower error component. The second subject may also have an error component of 1. The concept of reliability is defined as the extent of individual differences in the
characteristic being measured and the percent to which this is attributed to chance error (Anastasi, 1982).

Reliability, in other words, is the consistency with which a data collection instrument measures whatever it is measuring. If a test is reliable it means the measurement is consistent from time to time and researcher to researcher (Cox & West, 1986).

Reliability in Activity of Daily Living Assessments

Reliability of activities of daily living assessments estimates the amount of differences in activity of daily living (ADL) functioning among people. ADL functioning refers to the level of independence a person can achieve in the general tasks for living (i.e., dressing, grooming, cooking, etc.). The importance of the different types of reliability depends on the purpose of the measure. For an evaluative measure, the observer and test-retest reliability are important. If a descriptive measure is required, then internal consistency and observer reliability are important to the clinician (Law & Letts, 1989).

Unfortunately, there is no standard of time for evaluating test-retest reliability. Theoretically, in order to reduce the amount of error by scorers and administration, as well as unassigned and ineradicable error, one has to administer the retest immediately using the same administrator and scorer (Mehrens & Ebel, 1967). With functional assessments, however, immediately retesting a client can cause other negative factors such as fatigue, anxiety, or lack of
motivation (Hemphill, 1982). Since the purpose of an assessment for an occupational therapist is evaluation for decision making in treatment planning, the documentation of improvement, change, or status, accuracy was most important (Hemphill, 1982). Currently, there have not been enough observer and test-retest reliability studies on activity of daily living skills assessments documented in the current literature (Law & Letts, 1989).

Statement of the Problem

The purpose of this study was to evaluate test-retest reliability of the Levels of Rehabilitation Scale IIB (Carey & Posavac, 1982).

Summary

The literature stated that reliability is an important concept for clinicians in the areas of assessments. Without reliability the clinician could not be sure that the assessment is measuring actual changes in the client or if it is measuring something else. In order for the clinician to make informed accurate decisions, the assessment tool must be reliable. Of the four main types of reliability, this paper will look at test-retest reliability.
CHAPTER II

LITERATURE REVIEW

This literature review assessed the reason for examining the reliability and validity of the activities of daily living (ADL) assessments currently in practice. There was a review of the four main types of reliability. The final section of this chapter will focus on a review of the time factor between the initial test and the retest. It then gives a short review of past test-retest research.

Currently in the literature there is an overabundance of ADL assessments. Law and Letts (1989) suggested that the number of ADL assessments should be limited and more validation studies occur with those assessments already developed. In the review of assessments, Law and Letts found that only a few of the top assessments have established observer and test-retest reliability. These are the Barthel Index (Granger, Albrecht, & Hamilton, 1979), the Klein Bell ADL Scale (Law & Usher, 1988) and the PULSES Profile (Granger et al., 1979). The LORS IIB scale is an assessment tool which measures the functional independence of patients through their score on activity of daily living skills. Although it does not have test-retest reliability, the LORS IIB was rated excellent in almost every other category. These categories include clinical utility, scale
construction, standardization (good for standardization studies), internal consistency, observer reliability, and content and construct validity (Law & Letts, 1989).

Reliability has three main areas of concern. These are internal consistency, alternate forms, and test-retest (Cox & West, 1986). There is also a fourth concept of inter-rater reliability.

Internal Consistency

Internal consistency, also called split-half reliability, is accomplished by estimation of the reliability of the full length test by the correlation of the test of half lengths. The test, however, must be divided in a way as to assure that the two halves resembled each other in a statistical way, as well as the more superficial levels (i.e., length). This method can also be expanded by dividing the test into smaller fractions, even down to single items (Guilford, 1954). The results of this method would then tell the homogeneity of the instrument (Cox & West, 1986).

Alternate Forms

Alternate forms, on the other hand, measures the equivalence of the psychological-measurement content of one form of the test to another. The closer together in time of the administration, the more accurate the coefficient is a measure of internal consistency (Guilford, 1954). The two alternate forms should be measuring the
content in the same way in order for this method to be measuring reliability (Cox & West, 1986).

Inter-rater Reliability

Inter-rater reliability is defined as the consistency between two raters on scoring the subjects. It is important that all raters apply the tool consistently and reliably, or the data have no meaning. This is very crucial when more than one therapist is going to record data on the same client (Payton, 1988).

Test-retest Reliability

Test-retest reliability is defined when the same measure is administered twice to the same subjects. These two scores are then correlated to assess the congruity of the results (Cox & West, 1986). This coefficient does not relate to the internal consistency of a test. The importance of this procedure is that it evaluated the stability of the test. It measures the dependability of a test over time. High reliability can show that the individual remained uniform, in spite of changes in the environment, in that particular area that the test is measuring. Low reliability means that the functions being tested fluctuated from time to time or that the assessment, as an instrument, is affected by other things that fluctuated (Guilford, 1954).
Test-retest Examination

In a review of other studies, there is much fluctuation of the time factor between the initial test and the second retest. There is no research standard for the amount of time between testing to establish a significant reliability. Logic dictates that the longer a researcher waited between testing, the higher the risk of confounding factors skewing the data.

In a review of some psychological studies, there is great variance between the amount of time allotted between the initial test and the final retest. For example, the Wechsler Adult Intelligence Scale (WAIS) (Ryan, Georgemiller, Geisser, & Randall, 1985) was repeatedly assessed for test-retest reliability using a variety of time between testing. Kendrick and Post (1969), had clients take the test three times in a six-week period. In a later study of the WAIS, researchers used twenty weeks (Matarazzo, Weins, Matarazzo, & Manaugh, 1973) and two years (Rosen, Sallings, Floor, & Nawakiwska, 1973) as interval periods and documenting at least .84 significance correlation in scores. Finally, in 1985, Ryan et al. once again evaluated the WAIS for test-retest reliability using random time intervals from two to 144 weeks between testing. In this study interestingly, Ryan et al. (1985) distinguished between two terms, psychometric test-retest reliability and clinical test-retest reliability. Ryan et al. defined psychometric test-retest reliability as being highly reflected by a sizable test-retest correlation, and clinical test-retest as the absence of meaningful change in the score.
from the initial to the retest. This study found the psychometric test-retest was sufficient, but the clinical showed that 86% changed scores.

The same seemingly random choice of time interval is pervasive throughout the psychological research literature. In a study of the Marlowe-Crown Social Desirability Scale, a self-report assessment, the researchers used a six-week interval between administration (Zook & Sipps, 1985). However, in a study of another self-report assessment, the Minnesota Multiphasic Personality Inventory, the researcher decided on an interval of three days (Russell, 1986). There is usually no explanation for this discrepancy in intervals. It seemed to be an almost random choice of time span.

In a review of a few ADL assessments, a similar inconsistency is apparent. Of the 13 main ADL assessments chosen by Law and Letts (1989), only the Barthel Index (Granger et al., 1979), the Klein-Bell ADL scale (Law & Usher, 1988), and the PULSES Profile (Granger et al., 1979) show adequate test-retest reliability. For the Barthel index, the researchers chose the admission time, discharge time, and a two-year follow-up interview as the time frame (Granger et al., 1979). Law and Usher (1988) assessed the Klein-Bell ADL scale by retesting the subjects after a week. In this study Law and Usher (1988) established a test-retest reliability of .98.

Summary

In the current literature there is an overabundance of ADL
assessments. Law and Letts (1989) suggest that the assessments currently available be validated and researched for their value in the field, and no more assessments be developed.

In the literature there is reported four main types of reliability. They are: internal consistency, alternate forms, inter-rater reliability, and test-retest reliability. Each of these were reviewed and defined.

In a review of various test-retest reliability research, the literature reveals that there is no standardized amount of time between the initial and retest dates. With the Wechsler Adult Intelligence Scale (Brown & May, 1979), many different test-retest studies were conducted using intervals ranging from two to 144 weeks.

This seemingly random interval between the initial evaluation and the retest date is also revealed in the literature on ADL assessments. In the few ADL assessments that have been tested, the researcher uses from a week to a random discharge date as the length of time. For this study, there is to be an interim time of two weeks between the original assessment and the final retest. The research in this area is unclear as to a standard length of time between testing in order to prove test-retest reliability. This time was chosen for convenience to the researcher.
CHAPTER III

METHODODOLOGY

Subjects

The population for this study was eight volunteer post-rehabilitation stroke clients. These clients were randomly selected from registered stroke clients in the Marion Spears Teaching Clinic, located at Western Michigan University's campus. All of these clients were at least six months post-rehabilitation. The clients ranged in age from 43 to 73, the mean age being 62.8. There were five males and three females.

Instrument

The Level of Rehabilitation Scale IIB (LORS IIB) assessment measured the functional independence of patients. The scale concentrates on four main areas of function, including activity of daily living (ADL), mobility, communication, and cognitive ability. The scale was based on a ratio scale from 1-4. This scale had an equal distance between ratings. A rating of "1" indicated 25% independence, "2" indicated 50%, "3" indicated 75%, and "4" indicated 100%. The behavioral acts that distinguish the level of functioning have been determined by a consensus of therapists (Carey & Posavac, 1988).

11
Procedure

The assessment was administered by occupational therapy students trained to administer the assessment by the LORS IIB Procedural Manual. The clients were asked to come to the clinic for a one-hour evaluation. The subjects were then informed of the procedure for the test and asked to sign informed consent forms (see Appendix D). Six of the eight clients were assessed individually using the same ADL equipment available in the Marion Spears Teaching Clinic. Of the other two clients, one was assessed at home and one at the nursing home where she resided. At the end of this first testing, the clients were asked to return two weeks later to be retested. One of the eight was unable to return on the exact two-week date, but was able to be retested two weeks and four days later.

Research Question

In order to show adequate test-retest reliability, the data from the initial testing were a correlation with \( p \leq .05 \) agreement with the retested score within each category. The first group of experimental questions, Subscale I-V dealt with agreement between the Activities of Daily Living Subscale: (a) Dressing, (b) Grooming, (c) Washing and Bathing, (d) Toileting, and (e) Feeding.

The second group of experimental questions dealt with the agreement between the mobility rating scale: Mobility. The third area of agreement was the Community scale: (a) Auditory comprehension, (b) Oral expression, (c) Gestural expression, (d) Reading
comprehension, and (e) Written expression. The last area dealt with agreement between Cognitive abilities: (a) Memory, and (b) Problem solving.

Data Analysis

All of the subjects in this study were at least six months post-stroke. This fact made them a stable population for the span of the two weeks involved in the evaluation. Therefore, between their initial evaluations and the follow-up retest only a correlation with \( p \leq .05 \) agreement was considered significant.

The LORS IIB assessment was comprised of ratio data in each of the five categories. Since the ratio scale was comprised of fewer than seven possibilities it was treated as ordinal data. A Spearman Correlation was used.

In the following areas, these results were computed:

Activities of Daily Living Subscale
1. Dressing \( \rho = 1.00 \) \( \text{sig} = .000 \)
2. Grooming \( \rho = .8924 \) \( \text{sig} = .001 \)
3. Washing and bathing \( \rho = .9934 \) \( \text{sig} = .000 \)
4. Toileting \( \rho = 1.00 \) \( \text{sig} = .000 \)
5. Feeding \( \rho = 1.00 \) \( \text{sig} = .000 \)

The second group of experimental questions dealt with the agreement between the mobility rating scale.

1. Mobility \( \rho = .7333 \) \( \text{sig} = .019 \)

The third area of agreement was the Communication Scale.
1. Auditory comprehension $\rho = 1.00$ $\sigma = .000$
2. Oral expression $\rho = 1.00$ $\sigma = .000$
3. Gestural expression $\rho = 1.00$ $\sigma = .000$
4. Reading comprehension $\rho = 1.000$ $\sigma = .000$
5. Written expression $\rho = 1.00$ $\sigma = .000$

The last area dealt with agreement between Cognitive abilities.

1. Memory $\rho = 1.00$ $\sigma = .000$
2. Problem solving $\rho = 1.00$ $\sigma = .000$

The results show that the LORS IIB has great test-retest reliability in every category.
CHAPTER IV

DISCUSSION

(Poswac, 1982)

The results showed that the LORS IIB had very high test-retest reliability in every subcategory. This showed that the test was stable over a two-week period for this population. It is important that any assessment tool be tested for reliability and validity before it is used in clinical settings. The LORS IIB showed that it consistently measured this population the same. There were many factors that may have influenced these results. First, there were only eight subjects participating in the study. This small sample size may have increased the possibility of type II error. Another factor was that the subjects were all post-rehabilitation and therefore more stable than those still in the rehabilitation stage.

Some other considerations would be to adjust certain subtests, such as the written expression, problem solving, and memory to compensate for aphasic patients. It is common to have aphasic patients when working with stroke victims.

Another consideration for future studies would be to have a longer period of time between the testing. The two-week interim used for this study could be extended to a realistic period for a rehabilitation stay. The clients could be retested four to six weeks after the initial testing.
One further recommendation would be to have more than one rater in order to show inter-rater reliability.
CHAPTER V

CONCLUSION

Testing ADL assessments for validity and reliability was a vital step in the validation of the occupational therapy field. In a review of the literature there was an obvious lack of standardization for ADL assessments. It is believed that further studies should be conducted using a larger sample size and with different adult populations.

This study found a significant test-retest reliability for the eight stroke clients participating. A Spearman Rank order test was conducted and in all of the categories the rho was equal or less than one at a .05 level.
Appendix A

Letter of Consent From Human Subjects
Institutional Review Board
Dear Colleague:

The attached memo from the Human Subjects Institutional Review Board in relation to your proposed research is an important document which should be retained by you with your other research records. It is your proof of clearance of your research.

If you are a student: You may be required to produce this document when you turn in your thesis/dissertation to the Graduate College. Please retain the original so that you can meet the requirements.

Sincerely,

Mary Anne Bunda
Chair
Appendix B

Permission To Use LORS IIB Assessment
6 November 1989

Katherine Murphy
1940 Howard, Apt 550
Kalamazoo, MI 49003

Dear Ms. Murphy:

I hope you did not dispair of hearing from me. I am delighted to learn that you are interested in working with LORS-II.

The delay in getting back to you involved a number of things including a fall my wife took in which she broke an arm, sprained a foot, and brused a knee. She is on the mend now, but she was not ambulatory for a while.

The manual is enclosed and I included some comments regarding the meaning of test-retest reliability with a measure such as the LORS. I would be happy to provide reactions to a proposal draft if you want them. I would also appreciate a report on your findings.

Last, I would appreciate a copy of the article from The Journal of Occupational Therapy since the library here does not receive that journal.

Sincerely,

Emil Y. Posavac, Ph.D.
Chairman and Professor of Psychology
Appendix C

Letter of Informed Consent
Dear Sir or Madame:

I am a graduate student in occupational therapy at Western Michigan University. I am conducting a study at the Marion Spears Clinic in order to examine reliability of an assessment called the Level of Rehabilitation Scale II. Reliability means the assessment is consistently measuring whatever it is supposed to measure regardless of who is taking the measurements.

For this study, you will be asked to perform activities of daily living such as dressing, bathing, and feeding, while being assessed by an occupational therapy student. The rater will be trained occupational therapy students. At a later date approximately two weeks later, you will be asked to be retested on the same skills. Each assessment will take approximately 15-20 minutes.

Your name will not be recorded for any reason, so no one will be able to identify you in any way. There are no special risks involved in participation in this study. You are free to stop participation at any time.

Feel free to ask any questions you may have. You may leave a message for me or my research advisor, Barbara Hemphill, at 387-3860.

Thank you.

Sincerely,

Katherine Murphy MS OTS

I have read and understand all of the above information. All my questions have been answered, and I give my consent for_________________________ to participate

______________________________  ____________________________
Signature                        Date
BIBLIOGRAPHY


