

INFORMATION TO USERS

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

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

- 1 The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
- 2 When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
- 3 When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
- 4 For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
- 5 Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

**University
Microfilms
International**
300 N. Zeeb Road
Ann Arbor, MI 48106

8521102

Levinowitz, Arthur

A CRITERION-RELATED VALIDITY STUDY OF THREE SETS OF RATING
SCALES USED FOR MEASURING AND EVALUATING THE INSTRUMENTAL
ACHIEVEMENT OF FIRST AND SECOND YEAR CLARINET STUDENTS

Temple University

PH.D. 1985

**University
Microfilms
International** 300 N Zeeb Road, Ann Arbor, MI 48106

A CRITERION-RELATED VALIDITY STUDY OF THREE SETS
OF RATING SCALES USED FOR MEASURING AND
EVALUATING THE INSTRUMENTAL ACHIEVEMENT OF FIRST
AND SECOND YEAR CLARINET STUDENTS

Submitted to the Temple University Graduate Board
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

By
Arthur Levinowitz

April 1985



TEMPLE UNIVERSITY GRADUATE BOARD

Title of Dissertation: A CRITERION-RELATED VALIDITY STUDY OF THREE SETS OF RATING SCALES USED FOR MEASURING AND EVALUATING THE INSTRUMENTAL ACHIEVEMENT OF FIRST AND SECOND YEAR CLARINET STUDENTS

Author: Arthur Levinowitz

Read and Approved by: [Signatures: Elee C. ... , John M. ... , ... , S. ... , ... , ...]

Date submitted to Graduate Board: April 29, 1985

Accepted by the Graduate Board of Temple University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Date 5/2/85 [Signature] (Dean of Graduate School)

ACKNOWLEDGMENTS

The writer is deeply grateful to Professor Edwin Gordon for his many years of immeasurable support and exceptional guidance, to Professor John Holahan, and to the other members of the dissertation committee: Professors Stimson Carrow, Roger Dean, Eve Meyer, and Alan Sockloff. Special thanks are extended to his parents and to his wife, Lili, for their continued confidence and encouragement.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

LIST OF TABLES. vi

Chapter

 One PURPOSE OF THE STUDY 1

 Introduction 1

 Problems 8

 Two RELATED STUDIES 10

 Introduction 10

 The Abeles Study 16

 A Comparison of the Abeles Study
 and the Present Study 23

 The Levinowitz Pilot Study 25

 Three DESIGN OF THE STUDY 40

 Population 40

 Procedures 40

 Analysis 48

 Four RESULTS AND INTERPRETATION 52

 Means, Standard Deviations, and
 Intercorrelations of the X-Judges'
 Ratings 52

 Means, Standard Deviations, and
 Correlations among the Y-Judges'
 Ratings 58

 Intercorrelations Among the Dimensions
 of each Rating Scale Set used by
 the X-Judges 60

 Multiple Regression Analyses 64

 Interpretation 66

 Five SUMMARY AND CONCLUSIONS 70

 Purpose and Problems 70

 Design and Analysis 70

 Results 72

 Conclusions 74

| | |
|------------------------|----|
| APPENDIX A | 75 |
| APPENDIX B | 76 |
| BIBLIOGRAPHY | 77 |

LIST OF TABLES

Chapter Two

| | | |
|----|---|----|
| 1. | Nine Rating Scales | 27 |
| 2. | Means and Standard Deviations of Nine Rating Scales for each Judge and for Combined Judges | 31 |
| 3. | Interjudge Reliability Estimates for Nine Judges and Two Judges for Nine Rating Scales, and Intrajudge Reliability Estimates for Nine Rating Scales | 33 |
| 4. | Correlations Among Ratings on the Nine Rating Scales for each Judge and for the Combined Judges | 34 |

Chapter Three

| | | |
|----|---|----|
| 1. | Three Rating Scale Sets | 42 |
| 2. | Four Rating Scales used by each X-Judge | 46 |
| 3. | Three Rating Scales used by the Y-Judges | 48 |
| 4. | Steps for the Hierarchical Inclusion for the Multiple Regression Analyses | 51 |

Chapter Four

| | | |
|----|---|----|
| 1. | Means and Standard Deviations for each X-Judge and Group of X-Judges on each Rating Scale and Rating Scale Set | 53 |
| 2. | Interjudge Reliability Estimates for Four Judges, Three Judges, Two Judges, and One Judge (Intrajudge Reliability) for each Group of X-Judges on each Rating Scale and Rating Scale Set | 56 |
| 3. | Means and Standard Deviations of each Y-Judge's Ratings and Combined Y-Judges' Ratings on each Rating Scale | 58 |

| | | |
|----|---|----|
| 4. | Correlations Between Ratings of Y-Judge One and Y-Judge Two for each Rating Scale | 59 |
| 5. | Intercorrelations Among the Dimensions for the Combined Judges for Rating Scale Set 1 | 61 |
| 6. | Intercorrelations Among the Dimensions for the Combined Judges for Rating Scale Set 2 | 61 |
| 7. | Intercorrelations Among the Dimensions for the Combined Judges of Group I for Rating Scale Set 3 | 63 |
| 8. | Intercorrelations Among the Dimensions for the Combined Judges of Group II for Rating Scale Set 3 | 63 |
| 9. | Multiple Regression Analyses | 65 |

CHAPTER ONE

PURPOSE OF THE STUDY

Introduction

Most music educators and music researchers agree that rating scales are the most appropriate method for measuring and evaluating music performance.¹ There is a lack of consensus, however, regarding the specific design, content, and scoring procedure of rating scales.

Colwell recommends that a rating scale should be constructed to measure and evaluate "a single aspect of performance or a single element of understanding."² For example, a rating scale may be constructed to measure and to evaluate rhythm achievement, and a different rating scale may be constructed to measure and to evaluate tonal achievement.

Gordon suggests that two or three rating scales be used to measure the same music performance. Moreover, in

¹The Watkins-Farnum Performance Scale (1942-1952) is the result of an early attempt to develop an instrumental performance test. It remains today the only standardized performance test. The scale, however, does not use rating scales to measure and evaluate the performances.

²Richard Colwell, The Evaluation of Music Teaching and Learning (Englewood Cliffs, NJ: Prentice-Hall, 1970), p. 112.

addition to having each rating scale measure a separate dimension of performance, the dimensions should be independent.³ Independent dimensions are conceptually unrelated, and therefore, one would expect low correlations among the obtained scores for the dimensions. For example, a low correlation between students' scores obtained from a tonal rating scale and students' scores obtained from a rhythm rating scale, used to measure the same performance, would indicate that the two dimensions have little in common. The two rating scales may be used to measure two different aspects of the same performance. The low correlation between the scores obtained on a tonal rating scale and the scores obtained on a rhythm rating scale offers indirect evidence of the content validity and the construct validity of the two scales. It would remain to be demonstrated, however, that the scores obtained from the two rating scales are in fact valid measures of students' tonal and rhythm performance achievement.

There are many types of rating scales. Some types that are popular today are described by Cooksey.⁴ They are the summated (Likert), equal-appearing intervals, and the cumulative (Guttman) scales. Those three scales were

³Edwin E. Gordon, Learning Sequences in Music: Skill, Content, and Patterns. A Music Learning Theory. (Chicago: G.I.A., 1984), pp. 266-271.

⁴John M. Cooksey, "Developing An Objective Approach to Evaluating Music Performance," Symposium in Music Education, ed. Richard Colwell (Urbana-Champaign: University of Illinois, 1983), pp. 197-229.

developed originally to measure attitudes. When a Likert scale is used, an individual responds to a set of statements. He is asked to react to each statement by checking one of the "scaled responses." Most often, the scaled responses are: strongly disagree, disagree, neither disagree nor agree, agree, and strongly agree. An equal-appearing interval scale consists of dichotomous "scaled statements," with each statement assigned a predetermined scale value. An individual either accepts or rejects each statement. A Guttman scale consists of a set of statements, is used to measure one dimension, and is cumulative. An individual who agrees with a strong statement will also agree with a milder statement, and an individual who disagrees with a mild statement will also disagree with a stronger statement.⁵

Likert scales have been constructed to be used to measure and evaluate music performance. Cooksey states:

The Likert scale . . . has proven to be an effective approach to music performance evaluation. . . . In the evaluation of music performance, a major breakthrough occurred when facet-factorial analysis was applied to the construction of a rating scale for clarinet performance.⁶

⁵For additional information about Likert, equal-appearing intervals, and Guttman scales see R. Likert, "A Technique for the Measurement of Attitudes," Archives of Psychology, No. 140 (1932); L.L. Thurstone, The Measurement of Values (Chicago: University of Chicago Press, 1959); L. Guttman, "A Basis for Scaling Qualitative Data," American Sociological Review, No. 9 (1944), pp. 139-150.

⁶Cooksey, pp. 211,215.

Abeles states:

The rating scale seems the most practical approach to music performance evaluation . . . the forms currently in use do not reflect the current "state of the art" of rating scale construction. . . . A better procedure for constructing music performance rating scales is to model them on the Likert scales . . . such music performance rating scales . . . have reported high reliability and validity coefficients.⁷

Abeles used a Likert scale in conjunction with facet-factorial analysis to derive the final form of his rating scale.⁸ His Clarinet Performance Rating Scale (CPRS) consists of six dimensions. They are: Interpretation, Tone, Rhythm-Continuity, Intonation, Tempo, and Articulation. Each dimension consists of five statements. After listening to a clarinet performance, judges use a Likert scale and respond to each of the thirty statements. For example, the first of five statements that constitute the Interpretation dimension is: "effective musical communication." Each judge is to react to that statement after listening to an individual performance. His response is either strongly disagree, disagree, neither disagree nor agree, agree, or strongly agree.

⁷Harold F. Abeles, Charles R. Hoffer, and Robert H. Klotman, Foundations of Music Education (New York: Schirmer Books, 1984), pp. 252-253.

⁸Harold F. Abeles, "An Application of the Facet-Factorial Approach to Scale Construction in the Development of a Rating Scale for Clarinet Music Performance" (unpublished Ph.D. Dissertation, University of Maryland, 1971).

Abeles claims high validity for the CPRS. Colwell, however, states:

A valid overall rating is obtained but if diagnostic information is desired or if the scale is to make other than gross discrimination, additional research is needed.⁹

At least two other music researchers have used the Likert scale in conjunction with facet-factorial analysis to derive the final forms of their rating scales. Cooksey used the procedure to construct a rating scale to evaluate high school choral performance and DCamp used the procedure to construct a rating scale to evaluate high school band performance.¹⁰

Two other types of rating scales in use today are the continuous rating scale and the additive rating scale. "Continuous" and "additive" are two terms coined by Gordon to describe rating scales that he constructs to measure and to evaluate music performance.¹¹ Both the continuous rating scale and the additive rating scale include criteria.

⁹Richard Colwell, "Evaluation in Music Education: Perspicacious or Peregrine," Symposium in Music Education, ed. Richard Colwell (Urbana, Champaign: University of Illinois, 1983), p. 176.

¹⁰John M. Cooksey, "An Application of the Facet-Factorial Approach to Scale Construction in the Development of a Rating Scale for High School Choral Music Performance" (unpublished Ed.D. Dissertation, University of Illinois, 1974); Charles Barton DCamp, "An Application of the Facet-Factorial Approach to Scale Construction in the Development of a Rating Scale for High School Band Performance" (unpublished Ph.D. Dissertation, University of Iowa, 1980).

¹¹Gordon, p. 269.

Each criterion represents a different attainment level of an objective.

In a continuous rating scale, a hierarchy is inherent in the ordering of the criteria. A student is not credited with achieving the second criterion unless the first criterion has been achieved; a student is not credited with achieving the third criterion unless the second and first criteria have been achieved, and so on. By making all criteria in the rating scale interdependent in terms of continuous difficulty, measurement is emphasized in the continuous rating scale rather than evaluation.¹²

An additive rating scale is different from a continuous rating scale, because a hierarchy is not inherent in the ordering of the criteria. An additive rating scale resembles a checklist.¹³ A student is assigned the score that coincides with the number of criteria achieved. Evaluation is emphasized in the additive rating scale.¹⁴

Most rating scales consist of either three, five, or seven criteria. Gordon recommends the use of a five-point rating scale (five criteria) rather than a three-point or seven-point rating scale because "fewer than five criteria tend to decrease the reliability and more than five criteria tend to decrease the validity of the rating scale."¹⁵

¹²Gordon, p. 269.

¹³Colwell, p. 111.

¹⁴Gordon, p. 269.

¹⁵Gordon, p.266.

In a recent study, Gordon evaluated the vocal and instrumental performances of students in the fourth grade. Three dimensions were evaluated: rhythmic, tonal, and expression.¹⁶ Two five-point continuous rating scales were constructed: one for the rhythm dimension and one for the tonal dimension. An additive scale was constructed to measure and to evaluate the expression dimension.

The comparative validity of continuous scales and additive scales is not discussed in any of Gordon's publications. Gordon states, however, that when a student is to be rated on a dimension which has a basis in skill or content learning sequence, a continuous rating scale should be employed. When a student is to be rated on a dimension which does not as yet have a basis in skill or content learning sequence, it may be advisable to use an additive rating scale.¹⁷

In another recent study, the reliabilities of continuous rating scales and additive rating scales were compared.¹⁸ The rating scales were used to evaluate the rhythmic achievement of beginning clarinet students. Using

¹⁶Edwin Gordon, "A Longitudinal Predictive Validity Study of the Intermediate Measures of Music Audiation," Council for Research in Music Education, Bulletin No. 78 (Spring 1984), pp. 1-23.

¹⁷Gordon, Learning Sequences, p. 269.

¹⁸Arthur Levinowitz, "The Comparative Effectiveness of Different Rating Methods for Measuring the Rhythmic Skills in a Melodic Context of Beginning Clarinet Students" (unpublished manuscript, Temple University, 1984).

both continuous and additive rating scales, nine judges rated thirty individual performances of the same etude. The additive rating scale consisted of criteria identical to the continuous rating scale. The only difference between the two rating scales was the scoring procedure. Interjudge reliability of the continuous rating scale was .86, and the interjudge reliability of the additive rating scale was .77. Both reliability estimates are substantial. Thus, it is reasonable to suggest that the decision to use a continuous rating scale or an additive rating scale is not a question only of reliability but also a question of validity. The purpose of this research was to compare the validity of different types of rating scales for measuring and evaluating instrumental music achievement.

Problems

The appropriateness of three sets of rating scales for measuring and evaluating the instrumental music achievement of first and second year clarinet students was investigated. The three sets of rating scales are presented below:

- Set 1 - Gordon -- Tonal Continuous Rating Scale
Rhythm Continuous Rating Scale
Expression (Additive) Rating Scale
- Set 2 - Gordon -- Tonal Additive Rating Scale
Rhythm Additive Rating Scale
Expression (Additive) Rating Scale
- Set 3 - Abeles -- CLARINET PERFORMANCE RATING SCALES:
Interpretation, Tone, Rhythm-Continuity
Intonation, Tempo, and Articulation

The problems of this study were:

1. To compare the criterion-related validity of the three sets of rating scales for measuring and evaluating the instrumental music achievement of first and second year clarinet students.

2. To compare the content validity and construct validity of the three sets of rating scales for measuring and evaluating the instrumental music achievement of first and second year clarinet students.