

THE CONSTRUCTION AND VALIDATION OF RATING SCALES
FOR THE OBJECTIVE MEASUREMENT OF FIVE DIMENSIONS
OF ACHIEVEMENT IN CHORAL MUSIC PERFORMANCE

Submitted to the Temple University Graduate Board
in partial fulfillment of the requirements
for the degree of
Doctor of Musical Arts
by
Michael Howard Larkin
April, 1985

TABLE OF CONTENTS

LIST OF TABLES.iii
ACKNOWLEDGMENTSiv
Chapter		
One	PURPOSE OF THE STUDY	1
	Introduction	1
	Problems.	7
Two	RELATED STUDIES	8
	Introduction.	8
	The Watkins-Farnum Performance Scale	11
	The Facet-Factorial Approach.	14
	Comparison of Related Studies with the Present Study.	20
Three	DESIGN OF THE STUDY.	22
	Construction of the Rating Scales. .22	
	Reliability and Preliminary Validity Investigations	25
	Analysis.	27
	Rating Scales	29
Four	RESULTS AND INTERPRETATION	34
	Interjudge Reliability.	34
	Intercorrelations among the Rating Scales	44
	Interpretations	49
Five	SUMMARY AND CONCLUSIONS.	53
	Purpose and Problems.	53
	Design and Analysis	53
	Results	56
	Conclusions	57
BIBLIOGRAPHY.		58

CHAPTER ONE

PURPOSE OF THE STUDY

Introduction

A concern among music educators, especially those who are also conductors, is the lack of clearly defined objectives for their conducting courses as well as for their ensembles. In addition to haphazard teaching, another result of the absence of such objectives is that measurement and evaluation are not undertaken.

Music educators often confuse the meaning of the words measurement and evaluation. Gordon writes:

One of the serious problems in music education is that although some teachers rarely, if ever, measure, they often evaluate student achievement and their own teaching effectiveness. Though evaluation includes subjective elements, it must be based on objectivity.

If one accepts the premise that educational measurement yields achievement scores and evaluation is the interpretation of those scores, then intelligent evaluation is difficult without objective measurement.

¹Edwin Gordon, Learning Sequences in Music, Skill, Content, and Patterns (Chicago: G.I.A., 1980), p. 250.

The development and implementation of successful teaching is given direction by the identification of measurable objectives in addition to the identification of the most efficient sequence in which students might attain those objectives. An efficient teacher uses scores derived from a measurement instrument for assistance in the evaluation of student achievement of course objectives. The most effective teaching techniques and methods are of questionable value without measurement and evaluation, because the time to proceed to the next objective might be little more than guesswork.²

Unfortunately, many musicians are of the opinion that because of its many subjective elements, a complex behavior such as music performance does not lend itself to objective measurement. Conductors often claim that attempts to identify and measure specific performance objectives preclude the possibility of listening musically. They believe that music must always be listened to as a gestalt; in an artistic performance the whole is different from the individual parts. The whole musical experience, however, is shaped by achievement of many specific objectives which must first be identified and

²Edwin E. Gordon, Learning Sequences, p. 251.

then properly taught during the preparation of the performance. Madsen and Madsen write:

The scientific study of musical performance demands analysis and experimental verification in order to know which specific behaviors should be shaped and the best methods by which this can be accomplished. Music as an art has nothing to lose from the application of the scientific method. Objectification should not be confused with insensitivity.³

The identification, measurement, and evaluation of achievement relative to specific dimensions of a musical performance does not guarantee that a performance will be musical. Perhaps a conductor's greatest challenge is to create a musical and expressive performance from the various dimensions of performance, after each has received attention during the preparation of the performance.

Many music educators, including Colwell,⁴

³Clifford K. Madsen and Charles H. Madsen, Jr. Experimental Research in Music, in the Contemporary Perspective in Music Education Series, ed. Charles Leonhard, (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970), p. 40.

⁴Richard Colwell, The Evaluation of Music Teaching and Learning, in the Contemporary Perspectives in Music Education Series, ed. Charles Leonhard, (Englewood, New Jersey: Prentice-Hall, Inc., 1970), p. 112.

Leonard and House,⁵ and Gordon,⁶ believe that a rating scale is the most appropriate instrument for the measurement of music performance achievement. The objective measurement of music performance achievement, however, has been virtually ignored by music education researchers. Since 1964, only 16 of more than 2000 completed music education doctoral dissertations have dealt with the measurement of music performance achievement.⁷ Cooksey's facet-factorial approach to high school performance rating is the only doctoral dissertation that specifically deals with the measurement of choral performance achievement.⁸ To the writer's knowledge, the Watkins-Farnum Performance Scale,⁹ which is for the measurement of instrumental

⁵Charles Leonard and Robert W. House, Foundations and Principles of Music Education, Second Edition (New York: McGraw-Hill Book Co., 1972), p. 408.

⁶Edwin E. Gordon, Learning Sequences pp. 261-262.

⁷Richard Colwell, "Evaluation in Music Education: Perspicacious or Peregrine" in Symposium in Music Education: A Festschrift for Charles Leonard, ed. Richard Colwell (Urbana-Champaign: University of Illinois, 1982), p. 161.

⁸John Cooksey, "An Application of the Facet-Factorial Approach to Scale Construction in Development of a Rating Scale for High School Choral Performance", Diss. University of Illinois 1974.

⁹John G. Watkins and Stephen E. Farnum, Watkins-Farnum Performance Scale (Winona, Minnesota: Hal Leonard Music Inc., 1954).

performance achievement and is not a rating scale, remains the only published test of music performance achievement of any type.

Lack of consensus with regard to what dimensions of a choral performance can and should be measured objectively is fundamental to the lack of research pertaining to the measurement of choral performance achievement. In the United States, at least six different current schools of thought influence choral singing through theory and practice.¹⁰ It would be difficult to include all of those divergent ideas in the content of a single set of rating scales. Conductors must decide whether the content of a given set of rating scales is consistent with their own philosophies of choral singing and teaching methods. Research that pertains to the measurement of choral performance achievement should give insight into common ideas and practices among choral conductors, and into the specific criteria that might best define various dimensions of choral singing. Most importantly, increased awareness of objective measurement and evaluation by choral conductors should help to define more clearly their own philosophies,

¹⁰For a detailed discussion, see Howard Swan, "The Development of a Choral Instrument," in Choral Conducting: A Symposium, eds. Harold Decker and Julius Herford (Englewood Cliffs: Prentice-Hall, Inc., 1973), pp. 4-55.

objectives and teaching methods. The objectives to be measured should be a microcosm of the teacher's teaching method. To try to measure what has not been taught or learned is futile.

No published choral performance rating scales are currently available for which experimental evidence of reliability and validity have been obtained. There are, however, several adjudication sheets and rating scales available, such as the NIMAC Adjudication Form for Choral Music.¹¹ Those rating instruments emphasize subjective evaluation, not measurement, and they are most often used in conjunction with choral festivals and contests. Specific criteria within dimensions are either nonexistent or they are so vaguely defined that an adjudicator's ratings can be little more than subjective evaluations. Because those rating scales do not have clearly defined sub-scales, it is unlikely that they can be valid for diagnostic purposes. Moreover, the current performance achievement of an ensemble should be measured and evaluated more often than during festivals or public performances. Periodic

¹¹"The Organization and Management of Interscholastic Music Activities", in The NIMAC Manual, (Washington D.C.: National Interscholastic Music Activities Commission of the Music Educators National Conference, 1963), p. 15.

evaluation of the relative strengths and weaknesses of an ensemble, followed by appropriate teaching, may result in improved teaching, conducting, and performance.

PURPOSE

The purpose of this study is to investigate the feasibility of objective, diagnostic measurement of achievement in choral music performance.

PROBLEMS

The problems of this study are:

1. to outline choral music performance criteria, and from those criteria construct five-point rating scales, one each for the following dimensions of achievement in choral music performance: tone quality, balance/blend, rhythmic precision, expression, and intonation.
2. to establish reliability and other aspects of preliminary validity for the five rating scales for use with choral music of the Renaissance, Baroque, Classical, and early twentieth-century styles.

CHAPTER TWO

RELATED STUDIES

Introduction

There are few studies that are related even indirectly to the measurement of achievement in choral music performance. Cooksey's facet-factorial approach for the evaluation of high school choral performance remains the only dissertation that specifically deals with the measurement of achievement in choral music performance.¹

Researchers who have investigated the measurement of music performance achievement have dealt primarily with instrumental performance achievement. DCamp² and Abeles³ applied the facet-factorial approach to the measurement of band and clarinet performance achievement, respectively. Although the Watkins-Farnum Performance Scale is the best known

¹John Cooksey, "An Application of the Facet-Factorial."

²Charles Barton DCamp, "An Application of the Facet-Factorial Approach in the Development of a Rating Scale for High School Band Performance," Diss. University of Iowa 1980.

³Harold F. Abeles, "An Application of the Facet-Factorial Approach to Scale Construction in the Development of a Rating Scale for Clarinet Performance," Diss. University of Maryland 1971.

instrumental achievement test, Fleury⁴, Mansur⁵, Owen⁶, Iltis⁷, and Bostrom⁸ have conducted studies that have dealt specifically with the measurement of instrumental performance achievement.

Most researchers who have investigated the measurement of vocal and choral performance achievement have dealt with sight singing, or the ability of conductors to detect errors in a performance while concurrently reading the musical score. Early sight-singing tests were written by Mosher⁹ and Hillbrand¹⁰. More recent sight-

⁴Robert M. Fleury, "Objective Measurement of Group Instrumental Performance Music," Diss. University of California at Los Angeles 1963.

⁵Paul M. Mansur, "An Objective Performance-Related Music Achievement Test," Diss. University of Oklahoma 1965.

⁶Curtis Dawson Owen, "A Study of Criteria for the Evaluation of Secondary School Instrumentalists When Auditioning for Festival Bands," Diss. East Texas State University 1969.

⁷John Leon Iltis, "The Construction and Validation of a Test to Measure the Ability of High School Students to Evaluate Musical Performance," Diss. Indiana University 1970.

⁸Sandra J. Bostrom, "The Establishment of Content Validity and Estimation of Reliability of a Scale Developed to Evaluate Piano Performance," Diss. University of Southern California 1976.

⁹Raymond M. Mosher, "A Study of the Group Method of Measurement of Sight Singing" (New York: Teachers College, Columbia University Bureau of Publications, No. 194), 1925.

¹⁰E.K.Hillbrand, "Sight Singing for Grades 4 to 6" (New York: World Book Company), 1923.

singing achievement tests have been written by Cooper¹¹ and Scofield.¹² Tests that measure the ability of a conductor to detect errors during a choral performance have been written by Hansen¹³, Gonzo¹⁴, and Weatherly¹⁵.

The Watkins-Farnum Performance Scale¹⁶ will be examined in detail, because it remains the only published music performance test of any type, and also because of its unique scoring procedure. The Cooksey study also will be examined in detail, because it is the only dissertaion that pertains specifically to

¹¹John J. Cooper. "The Development of a Sight-Singing Achievement Test for Use with College Students," Diss. University of Kansas 1954.

¹²Willian Scofield. "The Construction and Validation of a Method for the Measurement of the Sight-Singing Abilities of High School and College Students," Diss. Michigan State University 1980.

¹³Louis A. Hansen. "A Study of the Ability of Musicians to Detect Melodic and Harmonic Errors in the Performance of Choral Music While Inspecting the Score," Diss. University of Kansas 1954.

¹⁴Carroll Lee Gonzo. "An Analysis of Factors Related to Choral Teacher's Ability to Detect Errors While Reading the Score," Journal of Research in Music Education, XIX, No. 3, (Fall 1971).

¹⁵Robert Weatherly. "Increasing Prospective Teachers Skill in Detection of Ensemble Performance Deficiencies by Means of Recorded Examples," Diss. Louisiana State University 1971.

¹⁶John G. Watkins and Stephen E. Farnum. Watkins-Farnum Performance Scale (Winona, Minnesota: Hal Leonard Music Inc., 1954).