Order Number 9411385

Development and validation of a method for diagnostic evaluation of solo musical instrument performance

Garczynski-Kessler, Ann Marie, Psy.D.

Pace University, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

DEVELOPMENT AND VALIDATION OF A METHOD FOR DIAGNOSTIC EVALUATION OF SOLO MUSICAL INSTRUMENT PERFORMANCE

by

Ann Garczynski-Kessler

A Doctoral Project Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Psychology in the Department of Psychology at Pace University

NEW YORK

1993

PACE UNIVERSITY

PSYCHOLOGY DEPARTMENT
PSY.D. PROJECT
FINAL APPROVAL FORM

(Please type all information)

NAME: Ann Garczynski Kessler

TITLE OF PROJECT: Development and Validation of a Method for Diagnostic Evaluation of Solo Musical Instrument Performance

DOCTORAL PROJECT COMMITTEE:

PROJECT ADVISOR: Dr. Alfred Ward
(Name)
Professor of Psychology, Pace University
(Title) (Affiliation)

PROJECT CONSULTANT: Dr. Barbara Mowder
(Name)
Professor of Psychology, Pace University
(Title) (Affiliation)

FINAL APPROVAL OF COMPLETED PROJECT:

I have read the final version of the doctoral project and certify that it meets the relevant requirements for the Psy.D. degree in School-Community Psychology.

Alfred W. Ward — 12/14/93
(Project Advisor's Signature) (Date)

Barbara G. Mowder — 12/14/93
(Project Consultant's Signature) (Date)

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

Many dear friends and colleagues contributed their time, skill, talent, ideas (and equipment) to make this project possible, including raters Rick McCurdy, Kristina Rizzo, and Steve Kessler; "model" pianists Karin McCartney and Joanna Raboy; and my friends Jeanne Porcino Dolamore, Paul Rivenburgh, Laura Gerak, Chris Gibbs, and Scott Hampton. My heartfelt thanks goes to each and every one of you.

In addition, I would like to thank the administrative people who supported this effort, namely Sheila Schwartz and Larry Balestra of the NYSSMA Executive Council, Larry Mullins of MENC, and the administration of the Sojourner Truth Library at the State University of New York at New Paltz (for extending borrowing privileges to a doctoral student who lived far from her own campus!).

But perhaps most of all, I wish to give thanks to those who consistently gave me something that I was not always able to hold on to, which was the belief in the value of my ideas and my ability to carry them out. I thank Dr. Bob Gibbs, who as my freshman advisor at the Crane School of Music encouraged me to think in terms of doctoral studies for the future. I thank my consultant, Dr. Barbara Mowder, who was always supportive and a pleasure to work with. And I owe special thanks to my advisor, Dr. Alfred Ward, whose enthusiasm and support for this project never waned, from the first discussions of an idea four years ago to the final presentation.

ABSTRACT

This research was designed to apply the school psychologist's expertise in the area of psychoeducational assessment to the development of a method of evaluating music performance achievement. Current methods are notoriously unreliable, and development of instruments with adequate psychometric properties has been hindered by beliefs that assessment of artistic endeavors must necessarily be subjective. However, consideration of research in the assessment of other school performance areas, especially oral reading, suggested that more objective measurement is possible. This project's goal was the development of a new measure that is reliable, valid, practical to use, and able to provide specific as well as general diagnostic information.

The author developed a two-part system for the evaluation of music performance achievement (SEMPA) to be applied to live solo performance. The first (or "objective") part of the SEMPA required coding of errors from portions of performances to be scored. The second (or "subjective") part was based on two highly reliable but obscure instruments that were modified for the SEMPA.

The research hypothesis of this study was that both parts of the SEMPA would prove to have adequate reliability, would demonstrate convergent validity, and would correlate with a criterion measure. Results of the study indicated that

interrater reliability and scoring accuracy was high for most categories, but reliability varied in the Pitch/Intonation category (reliability was higher for passages identified as more difficult to score). Convergent validity correlations ranged from .35 to .73 for Part I scores when adjusted for error equivalence. Both SEMPA Part I and Part II total ratings correlated with the criterion measure (correlations ranged from .61 to .82).

The study provided support for the SEMPA as a reliable and valid method for evaluation of live student performance. However, the study raises some questions about rater ability to perceive errors and other issues that can be addressed in future research. Although the SEMPA was originally inspired by standardized oral reading tests, the process of its development and its final form very closely parallels that of Curriculum-Based Measurement (CBM) systems. A number of considerations from that literature have implications for future research with the SEMPA.

Chapter I - Introduction

Achievement tests have continued to be of great importance in the schools, and surpass all other types of tests in terms of sheer numbers (Anastasi, 1988). However, it is only recently that interest has been focused on the measurement of achievement in music, and there are relatively fewer music achievement tests in existence when compared with other school subject areas.

The December 1989 issue of the *Music Educators Journal* contains a special focus on musical evaluation. One contributor comments on the continuing demand for accountability and objective evidence of student achievement. However, tests of musical achievement and related constructs have been subject to criticism which centers around two points: 1) tests are inadequate and unfair, and 2) test results are misunderstood and misused.

The first criticism is aimed at the tests themselves and suggests that test development needs to be a more rigorous process than it has been in the past. The second criticism addresses the training and level of competence of those who administer and make decisions based on test results. Both of these issues have been addressed in recent years by researchers in music education who are aware of the need for accurate instruments for aptitude and achievement

assessment, diagnostic assessment, and program evaluation (e.g., Abeles, 1973; Gordon, 1967; 1975).

A particularly difficult problem with assessment in music education occurs when students who receive instruction in playing a musical instrument must be evaluated to determine their level of achievement. Presently, most teachers are left to develop their own method of evaluation, for there is no one universally accepted system or method of performance achievement evaluation. This reliance on teacher-made methods does not pose as great a problem when the same teacher assesses a group of students whom he or she sees on a regular basis. In fact, there are those who hold that assessment of the creative arts must necessarily be subjective, and dependent on intuitive assessment (Amabile, 1983). It is when the achievement of a student must be evaluated in relation to other students, who have other teachers, that the problem of not having an objective and appropriate method of evaluation becomes serious.

The problem of evaluation of performance achievement by raters other than the student's own instructor is important to address for at least two reasons. One is that musicians who play in ensembles such as a band or an orchestra traditionally compete for their seating within their section, and for admission into more select ensembles. In a research context, the ability to measure music performance achievement is crucial in the development and validation of

effective teaching methods. Unfortunately, many researchers consider the problem unsolvable because of the complexity of assessment in the arts.

A review of currently available tests and/or systems of evaluation of music performance suggests that the criticisms mentioned previously (inadequacy and misuse) are warranted. Subjective systems are the norm, and more objective systems (e.g., Farnum, 1969; Pizer, 1987) are developed intuitively, are not validated, and/or are impractical to use. Few methods have been based on previous research. However, consideration of the research in the assessment of another school performance area, oral reading, suggests objective measurement is not impossible. In the area of oral reading, standardized and highly objective measures have been constructed and used successfully in the assessment of achievement in an area where the evaluator must listen to an individual performance. These measures have been able to provide a wealth of diagnostic information as well.

School psychologists have traditionally been interested in measurement issues and facilitation of the instructional and learning process, and expertise in individual assessment continues to be an important area of competence for the school psychologist (Marston, 1989). Because school psychologists work directly with educators and students in schools, they have been able to apply their understanding of the constraints of the instructional setting and the need for

objective evaluation to the development of improved instruments, and have been particularly successful with individualized assessment. An example of a recent contribution by school psychologists is the development of measurement systems based on actual curriculum materials, rather than measures using standardized test items or passages ("curriculum-based measurement" or CBM) (Marston, 1989). Although music is a school performance area that has not received a lot of attention in the school psychology literature to date, it appears that a school psychologist is in an ideal position to assist with developing solutions to problems with objectivity and measurement in any subject area, and the problems identified in the assessment of music performance are by no means unsolvable when viewed from the school psychologist's perspective.

The purpose of this research was to develop a method of music performance evaluation that is valid, that meets high standards of reliability, and that is as practical to use as the most commonly used systems today. This was done through the application of psychometric principles, guided by past research in oral reading evaluation, to the musical audition. In addition to the method itself, the development of evaluator training procedures that will help insure standardization of administration was of central importance.

A Brief History of Measurement and Evaluation of Musical Behaviors

The earliest efforts at developing measures of musical constructs were aimed at assessment of music aptitude. Interest in the measurement of music aptitude, which is defined as potential or capacity for achievement (Lehman, 1968, p. 8), started growing in the early 1900's. In 1919, psychologist Carl Seashore published the first battery to gain wide recognition. Over the next 40 years, Seashore revised and refined his measure, and the latest revision is still available for use today (Seashore, Lewis, & Saetveit, 1960). An eminent British music psychologist named Wing found a general factor of musical ability that he believed pervaded musical ability. His research resulted in the *Standardised Tests of Musical Intelligence* (Wing, 1961). These tests, which reflect the importance of a general factor of musical ability, are in sharp contrast to Seashore's measure, which involves a specifics approach.

In the years before 1960, there was relatively little interest in the development of measures of musical constructs. The majority of tests developed during this period were paper-and-pencil tests of music achievement and aptitude. Measurement and evaluation of music performance achievement took place in the form of auditions and contest ratings, but there was no formal attention by researchers to anything but the psychophysical elements of music performance.

A surge of interest in formal means of measurement of musical behaviors occurred in the 1960's. Three books from that period were important in the advancement of the field of music and measurement: Whybrew's *Measurement and Evaluation in Music* (1962), Lehman's *Tests and Measurements in Music* (1968), and Colwell's *The Evaluation of Music Teaching and Learning* (1970). This surge of interest coincided with the growth of the accountability in education movement in the 1960's and 70's, and education in the arts was closely scrutinized during this time of economic belt-tightening. This period saw the introduction of many new methods for evaluation of music instruction, musical attitudes and other affective variables. Evaluation of whole programs were developed, as well as tests of musical aptitude, achievement, and, for the first time, achievement as demonstrated through actual performance.

Parallel: A Brief Survey of Issues in the Measurement and Evaluation in Reading

In comparison to the field of measurement and evaluation of other educationally related constructs, the field of measurement in music is quite young. It is fairly clear that the reasons for this lag in development have to do with the demand–before 1960, there was no pressure to demonstrate quantifiable results in any musical endeavor, and subjectivity in measurement was generally unquestioned. However, such was not the case with reading education. The necessity of reading ability in order to be successful in life was

established in the minds of most people in the 1800's, and there was pressure to demonstrate the success of particular methods of reading instruction. In 1894, Rice developed the notion of assessing instructional effectiveness with specific tests, and by 1914 his ideas gained such acceptance that large scale surveys of educational progress in the core areas were well underway (Johnston, 1984).

Of the many measures of reading achievement in use in the early 1900's, the measures that became most popular were the ones that were the most efficient: those that involved silent reading and the solving of text related questions. The history of such assessment dates back to the time of Thorndike, who is considered the father of modern group reading tests as well as the father of modern reliability theory. The assessment of comprehension was the main focus, and Johnston (1984) states that in contrast to the development of silent reading assessment, the development of oral reading assessment has been "painfully slow." Criticisms of the use of oral reading in assessment were that it did not deal with understanding, and was not the most important kind of reading. Oral reading was used in the in assessment process, but drew little comment in the research literature in the first two decades of the 20th century. Gray's *Standard Oral Reading Paragraphs* were introduced in 1915. Gray's test consisted of a graded series of paragraphs read aloud. The examiner coded errors such as omissions, additions, substitutions, mispronunciations, etc., and

norms were provided for reading time. In 1937, Gray refined the recording procedures and error types.

Gray's method of recording errors in oral reading was largely ignored. Descriptive and comparative studies of errors increased in frequency, but there were few attempts to use errors to model the processes, knowledge, and misunderstandings that produced them until the 1970 publication of *Reading Miscue Analysis* by Goodman and Burke.

Johnston (1984) notes that informal oral reading tests were probably downplayed throughout the history of reading assessment because of their lack of objectivity, and also due to an ideological thrust that favored ease of use over all else. However, in the last 20 years, the value of the oral reading assessment in the measurement and evaluation of overall reading ability has been acknowledged, and there have been advances in the psychometric properties of formal and informal oral reading tests. Issues of concern have been reliability, validity, readability and passage equivalence, objectivity, efficiency, the inference of underlying processes based on error patterns and the process versus product model, and the importance of teacher training in the assessment process.

The history of the development of the assessment of oral reading has many parallels to the history of the development of the assessment of music performance. Both oral reading and music performance are a subset of larger