

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

AN INVESTIGATION OF THE ADEQUACY OF THE
CONTENT AND DIFFICULTY LEVELS OF THE
IOWA TESTS OF MUSIC LITERACY

This is to certify that the thesis of
Warren C. Swindell with a major in Music has been
approved by the Examining Committee as satisfactory
for the thesis requirement for the Ph.D. degree at
the convocation of August, 1970.

Thesis committee:

Edwin Gordon
Thesis Supervisor

Darrell Sabers
Thesis Co-Supervisor

by

Warren C. Swindell

W. H. Johnson
Member

H. H. Johnson
Member

A thesis submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the School of Music in the Graduate College
of The University of Iowa

August, 1970

Edwin Gordon
Member

Thesis Supervisor: Professor Edwin Gordon
Co-Supervisor: Asst. Professor Darrell L. Sabers

TABLE OF CONTENTS (continued)

Chapter	Page
TABLE OF CONTENTS	
V. SUMMARY AND CONCLUSIONS	131
Summary	131
Results	133
Conclusions	137
Recommendations for Further Study	139
Chapter	Page
I. PURPOSE OF THE STUDY.	1
Introduction	1
Iowa Tests of Music Literacy	19
Purpose of the Study	20
Problem of the Study	21
Limitations of the Study	30
II. A COMPARISON OF MUSICAL ACHIEVEMENT TESTS AND IOWA TESTS OF MUSIC LITERACY.	33
Introduction	33
Rationale and Purpose of the Batteries	34
Content of the Tests	39
Experimental Validity of MAT and ITML.	47
Provision for Assessing Levels of Growth	48
Reliability of the Batteries	51
III. PROCEDURES.	55
Introduction	55
Description of the Sample.	57
Administration of the Tests.	58
Design and Analysis.	63
IV. PRESENTATION AND INTERPRETATION OF DATA	66
Introduction	66
Data Within Grades	66
Data Among Grades.	84
Test Intercorrelations	92

TABLE OF CONTENTS (continued)

Chapter		Page
V.	SUMMARY AND CONCLUSIONS	131
Table	Summary	131
	Results	133
1.	Conclusions	137
	Recommendations for Further Study	139
	BIBLIOGRAPHY	142
	APPENDICES	147
	Appendix A	148
	Appendix B	162
	Appendix C	169
	3. Total Number of Examinees by School and Level of ITMC	59
	4. Maximum Correct Scores for Each ITMC Subtest, Total Test, and Composite Test	67
	5. Fourth Grade Means, Standard Deviations, and Reliabilities for Levels 1 Through 6	68
	6. Fifth Grade Means, Standard Deviations, and Reliabilities for Levels 1 Through 6	69
	7. Sixth Grade Means, Standard Deviations, and Reliabilities for Levels 1 Through 6	70
	8. Seventh Grade Means, Standard Deviations, and Reliabilities for Levels 1 Through 6	71
	9. Eighth Grade Means, Standard Deviations, and Reliabilities for Levels 1 Through 6	72
	10. Ninth Grade Means, Standard Deviations, and Reliabilities for Levels 1 Through 6	73
	11. Level 1 Means, Standard Deviations, and Reliabilities for Grades 4 Through 9	85
	12. Level 2 Means, Standard Deviations, and Reliabilities for Grades 4 Through 9	86

trial and error; the psychologists have for a comparatively short period conducted systematic experiments."

Chapter I

PURPOSE OF THE STUDY

Ross, after tracing the history of various tests and measurements back to Antiquity, concluded that

Introduction

The late eighteenth century is described by Grout as a milestone in Western Civilization. At that point, man increasingly began to develop faith in the efficacy of applied experimental knowledge.¹ Much of man's progress can be, indeed, attributed to adherence to scientific methodology. Although musician-educators such as Rousseau were at the forefront of the movement to use objective methods as the starting point of investigation, it is somewhat paradoxical that trial and error methods continued to be used to solve problems in music education. Not until the relatively recent emergence of psychology did man attempt to solve problems in music education by means of systematic experimentation. Wing has succinctly assessed the dilemma: "The musicians have worked for centuries by

¹Donald Jay Grout, A History of Western Music (New York: W. W. Norton & Co., Inc., 1960), pp. 411-412.

trial and error; the psychologists have for a comparatively short period conducted systematic experiments."²

Early Measurement Attempts

Ross, after tracing the history of various tests and measurements back to Biblical Antiquity, concluded that more progress had been made between 1916 and 1941 than all preceding years.³ Thus, testing has been a traditional part of teaching processes; but where teaching was regarded as a highly skilled art, testing was regarded as something that anyone could do quite casually. Horace Mann, for example, wrote about the importance of testing, over one hundred years ago, and contributed to the demise of oral examinations when he clearly demonstrated the superiority of written tests.⁴ In most situations, written

²Herbert Wing, "Tests of Musical Ability and Appreciation; an Investigation into the Measurement, Distribution and Development of Musical Capacity," The British Journal of Psychology; Monograph Supplements, XXVII, p. 1.

³Clay C. Ross, revised by Julian C. Stanley, Measurement in Today's Schools, third edition (New York: Prentice-Hall, Inc., 1954), p. vii.

⁴Otis W. Caldwell and Stuart A. Courtis, Then and Now in Education: 1845-1923 (Yonkers: World Book Co., 1923), pp. 37-41.

examinations were more valid, reliable, and usable⁵ than oral types which were then used almost exclusively.

Several studies conducted during the second and third decades of this century collectively and conclusively stressed the inherent unreliability of school grades and examinations. One study of grades completed at the University of Missouri revealed astonishing variations in the distribution of class grades. For example, 55% A's were awarded in philosophy but only 1% in Chemistry III. There were 28% failures in English II and none in Latin I. A similar study conducted over a two year period at the University of Chicago High School disclosed 17.1% A grades and 8.4% failures for German classes, whereas English teachers recorded only 6.5% A's and 15.5% F's. The following conclusion was logical:

Such variations both at Chicago and Missouri could be most reasonably interpreted on the supposition, not that English is harder than foreign languages, but that English instructors are harder. In other words, school marks are highly subjective, the

⁵Usability is defined by Ross as practicability. "By this is meant the degree to which the test or other instrument can be successfully employed by classroom teachers and school administrators without an undue expenditure of time and energy . . ." Ross, Measurement in Today's Schools, p. 127.

The mark received often being more a function of the personality of the instructor than of the performance of the student.⁶

Other data were even more damaging. Facsimiles of the same geometry paper were marked by 116 mathematics teachers. Grades assigned to the paper ranged from 28 to 92. Ironically, mathematics is one of the most objective subjects. Another study of grades awarded by English teachers was even more spectacular. One hundred English teachers were asked to evaluate a composition by assigning it a percentage value and to indicate the school grade in which they would expect that quality of work to be done. Assigned percentage values ranged from 60 to 98, while the estimated grade location ranged from fifth grade to the junior year of college.⁷ Professional educators were appalled at the lack of well-ordered grading procedures.

⁶Ibid., p. 39.

⁷The composition had actually been written by a talented high school senior whose special interest was journalism and who also did correspondence work for some Chicago newspapers. A few years earlier it had been deemed the best found by a survey committee at Gary, Indiana. Ibid., pp. 40-41.

Proceedings of the National Education Association, LXXIX, 1941, pp. 323-344.

Thorndike,⁸ to be sure, considered such subjective practices scandalous.

Standardized Achievement Tests

The movement to measure educational attainment quantitatively was capably led by Edward L. Thorndike. His dictum, which follows, is now illustrious: "Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality."⁹ One year later, test making had begun to reach a more sophisticated level.¹⁰

Early standardized achievement tests were usually of the general or survey type. Such tests afforded a general assessment of the pupil's attainment in a subject, but did not provide the detailed information required for

⁸Edward L. Thorndike, "The Nature, Purposes, and General Methods of Measurements of Educational Products," The Measurement of Educational Products; Seventeenth Year-book of the National Society for the Study of Education (Chicago: National Society for the Study of Education, 1918), pp. 16-24.

⁹Ibid., p. 16.

¹⁰R. B. Buckingham, "Our First Twenty-five Years," Proceedings of the National Education Association, LXXIX, 1941, pp. 323-344.

remedial work. Diagnostic tests appeared next; these instruments provided specific information regarding a pupil's strong and weak points. The organization of tests into batteries made up of survey tests in core subjects, all published in a single booklet, was a major innovation. The first edition of The Stanford Achievement Test appeared in 1922. Nevertheless, widespread acceptance of standardized tests was slow because they represented a considerable item of expense which school boards during the 1920's were often reluctant to assume.¹¹

Musical Achievement Tests

Musical achievement tests emerged about the time standardized tests were beginning to reach more refined levels. In spite of fundamental differences in the nature of music,¹² when compared with subjects like mathematics,

¹¹Ross, Measurement in Today's Schools, p. 44.

¹²More recently, the nature of music and fundamental objective for music education have been clearly enumerated by Gordon: "A human infers musical meaning from musical sound because he is able to organize, and therefore understand, what he hears. . . . In that music is an aural art, one must acquire musically meaningful aural and kinesthetic reactions for the purpose of developing music appreciation." Edwin Gordon, Psychology of Music Teaching (To be published by Prentice-Hall), p. 70.

basic sciences, and English, early musical achievement tests were patterned after instruments in other fields. Initial musical achievement tests measured factual learning such as knowledge of musical symbols, terms, key and meter signatures, interval structure, and similar rudiments of music theory. Few tests measured a subject's ability to associate what is heard with what is seen in notation.

For example, The Beach Music Test consists of eleven parts and deals mostly with factual knowledge. First published in 1920, it is now of only historical importance. The Kwalwasser-Ruch Test of Musical Accomplishment was published in 1925. Ten subtests cover the following areas: (1) Knowledge of Musical Symbols and Terms, (2) Recognition of Syllable Names, (3) Detection of Pitch Errors in a Familiar Melody, (4) Detection of Time Errors in a Familiar Melody, (5) Recognition of Pitch Names, (6) Knowledge of Time Signatures, (7) Knowledge of Key Signatures, (8) Knowledge of Note Values, (9) Knowledge of Rest Values, and (10) Recognition of Familiar Melodies from Notation. Because musical accomplishment includes more than familiarity with notation, the test title is, at best, misleading. For instance, no attempt was made to

measure creativity, nor performance ability, which are two very important aspects of musical accomplishment. Therefore, other than apparent face validity as a test of musical notation, the instrument is not an adequate measure of musical accomplishment. The Kwalwasser Test of Music Information and Appreciation deals with factual knowledge such as music history, biography, instrumentation, and form. Unfortunately, Kwalwasser failed to report reliability and validity data. Published in 1927, much of the test content is now outdated.

It was not long before responsible professional educators began to criticize achievement tests which emphasized rote learning and factual knowledge. E. F. Lindquist, who has made vast contributions to the field of educational measurement, was justly critical of instruments which failed to measure understandings. His assessment of early trends was important:

It is unfortunately true that a large share of present-day instruction at the secondary school level, and in the college as well, is characterized by close dependence upon the textbook, by much "lesson learning" of the memory type, and by mechanical drill and "recitation" procedures. Such instruction has placed a distinct premium upon rote learning of statements, facts and of

principles; upon conscientious memorization of the unique phrasing of the testbook or of the lecture "notes"; and upon conformity with the opinions of the textbook author or of the instructor. It is not surprising, therefore, to find that the tests which have been used in connection with such instruction have been almost "informational" in character, in the sense that they have failed to measure the student's real understanding of the subject or his ability to do inferential thinking in the field to be tested.¹³

It is truly deplorable that 33 years after Lindquist's sweeping indictments many music teachers are still using the memory approach to teaching and testing.

Knuth was one of the first musical achievement test authors to focus upon understandings rather than the learning of rigid answers to rigid questions. The Knuth Achievement Tests in Music (forms A and B) are, in fact, sometimes regarded as two of the oldest instruments which are still feasible for use. Researchers have used the tests to compare recent musical achievement with achievement more than two decades ago. Although first published in 1936, the tests were revised in 1966. The ability to associate

¹³American Council on Education, Herbert E. Hawkes, E. F. Lindquist, and C. R. Mann (editors). The Construction and Use of Achievement Examinations; A Manual for Secondary School Teachers (Boston: Houghton Mifflin Co., 1936), p. 82.

musical notation with aural stimuli is the only aspect of musical achievement actually measured. The Strouse Music Test, published in 1937, has perhaps inhibited rather than advanced the musical achievement test movement. Factual in nature, much of the test is faulty because of the arbitrary and ambiguous nature of the items. Such tests evoked more skepticism among a large core of music teachers who were already dubious about the merits of published batteries.¹⁴

Constructed during the forties, The Jones Music Recognition Test consists of matching performed melodies with musical notation. However, the test is unstandardized and neither validity nor reliability data are available. The Farnum Music Notation Test was designed for grades 7, 8, and 9. Its date of publication was 1953. As with the Jones Tests, subjects are asked to indicate whether

¹⁴McConn explains that teachers were bewildered by the entry of objective achievement tests. Such tests were looked upon as a series of strange new inventions, nearly all of them appallingly elaborate, and alleged to have been laboriously prepared, with every item studied and checked, and then re-studied and re-checked by mysterious statistical methods. See: Max McConn, "Examinations Old and New; Their Uses and Abuses," The Educational Record XVI/4, 1935, pp. 375-376.