

University of Montana

ScholarWorks at University of Montana

Graduate Student Theses, Dissertations, &
Professional Papers

Graduate School

1967

Investigation into certain aspects of evaluation as they pertain to senior high school music education

William Robert Swaffield
The University of Montana

Follow this and additional works at: <https://scholarworks.umt.edu/etd>

Let us know how access to this document benefits you.

Recommended Citation

Swaffield, William Robert, "Investigation into certain aspects of evaluation as they pertain to senior high school music education" (1967). *Graduate Student Theses, Dissertations, & Professional Papers*. 2742. <https://scholarworks.umt.edu/etd/2742>

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.

AN INVESTIGATION INTO CERTAIN ASPECTS OF EVALUATION
AS THEY PERTAIN TO SENIOR HIGH SCHOOL MUSIC EDUCATION

By

William Robert Swaffield

B.A. University of Saskatchewan, 1956

B.Ed. University of Alberta, 1958

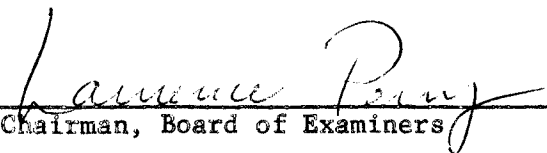
Presented in partial fulfillment of the requirements for the degree of

Master of Music

UNIVERSITY OF MONTANA

1967

Approved by:


Chairman, Board of Examiners


Dean, Graduate School

OCT 13 1967

Date

UMI Number: EP34847

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



UMI EP34847

Published by ProQuest LLC (2012). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to Dr. Gerald Doty, not only for his encouragement, but also for the extra time which he spent in counsel relative to this paper.

The author also wishes to thank Mr. Laurence Perry and Dr. Lloyd Oakland for their guidance and help.

TABLE OF CONTENTS

CHAPTER	PAGE
I. The Problem	1
Introduction.	1
Statement of the Problem.	1
Delimitations of the Study.	2
Definitions	2
Basic Assumptions	8
Purpose of the Study.	8
Method of Procedure	9
Review of the Literature.	10
Brief Historical Background of Evaluation in General.	18
II. The Nature of Evaluation in Music Education	19
Purposes of Evaluation in Music Education	26
Limitations of Evaluation in Music Education.	29
III. Two General Methods of Evaluation in Music Education.	33
Formal.	33
Informal.	34
IV. Two Specific Types of Tests and Testing Procedures.	36
Objective	36
Subjective.	39
V. Suggested Methods of Evaluation and Specific Types of Tests and Testing Procedures in Three Areas of High School Music Education	43
General	43
Choral.	45
Instrumental.	48
Conclusions	50
Recommendations	51
Bibliography.	52
Appendixes.	56

CHAPTER I

The Problem

Introduction

It is common for high school students in any subject area to wonder, often audibly, how grades are determined. Unfortunately, this question has long been a notoriously embarrassing one for most music teachers, simply because it requires the existence both of clearly defined standards of comparison and objectives of study.

Hopefully, the situation has improved in recent years. In cases, however, where the teacher is dissatisfied with his evaluative procedure, this study may be of some assistance.

Often the problem springs from inadequate or even non-existent instruction in the available materials and techniques of evaluation at the teacher education level. Such a situation should not exist in the face of research that has been conducted and the many tests and testing procedures which have been developed.

Statement of the Problem

1. What is the nature of evaluation as it relates to music education at the high school level?
 - a. Why should the teacher evaluate his students?
 - b. What evaluative difficulties arise from the nature of music itself?
To what extent does the nature of test construction and grading influence the quality of evaluation?

2. What general methods of evaluation are available to the music teacher?
3. What specific types of tests and testing techniques are available to the teacher?
4. Which tests or testing techniques would be most effective in certain specified areas of music study?
 - a. General music.
 - b. Choral music.
 - c. Instrumental music.

Delimitations of the Study

The study is limited to evaluation in three areas of senior high school music education: general, choral, and instrumental music (band or orchestra). It does not include, for instance, suggestions regarding private music study. It is also limited to concepts, materials, and techniques which the author believes to be effective, not necessarily those which are actually being used.

Definitions

Ability - The power to perform a designated responsive act.

Accuracy - In testing, the quality of pinpointing specifically and dependably what the student knows or can do and vice versa.

Achievement - A term which refers to the level of proficiency and understanding which a student has reached as a result of instruction.

Administration - In testing, a term which refers to the method which is used to present the test to the student.

Appreciation - A term which refers to the degree of understanding and resulting enjoyment which a student possesses of a work of art.

Aptitude - The capacity to acquire proficiency with training.

Aptitude Test - A test which measures this capacity.

Achievement Test - A test that measures the amount a pupil has achieved in subject areas or in general schooling.

Capacity - A term which refers to maximum ability with further training, or potential ability.

Comprehension - In testing, the quality of covering the range of learnings in a particular area being tested.

Correlation Coefficient - This is the most commonly used measure of relationship between paired facts or of the tendency of two or more variables or attributes to go hand-in-hand. It ranges in value from -1.00 for perfect negative relationship through 0.00 for none or pure chance to +1.00 for perfect positive relationship.

Criterion - A standard, norm, or judgement used as a basis for quantitative and qualitative comparison.

Cumulative Record - A collection of information about a student covering a number of school years and including such factors as academic marks, and information about health, family, interests, school success, aptitudes, and social adjustment.

Diagnostic Test - A test whose purpose it is to determine a student's strengths and weaknesses in particular subject areas.

Discriminating Power - The ability of a test item to differentiate between individuals possessing much of some characteristic (skill, knowledge, aptitude) from those possessing little of this characteristic.

Essay Test - (Also called "discussion question" test.) A test in which the student is asked to express his understandings of a subject. This type of testing device is particularly useful in measuring non-factual factors such as attitudes, interests, creativity, and verbal expression. It takes the forms of extended-response, restricted response, definition, and outline.

Evaluation - A judgment as to how close a student has come to desired behavior in terms of clearly defined objectives. (Broader than measurement)

Evaluation Program - The testing, measuring and appraisal of the growth, adjustment, and achievement of the learner by means of tests and non-test instruments and techniques. It involves the formulation of objectives, their definition in terms of pupil behavior, and the selection or construction of valid, reliable, and practical appraising instruments.

Formal Evaluation - Methods of judging student progress which are standardized and meet acceptable levels of reliability and validity.

General Music Programs - A course of study in music which attempts to give the student wide experience in the world of music through the consideration of (1) music history and literature, (2) theory, including sight-singing and keyboard skills, and (3) vocal production.

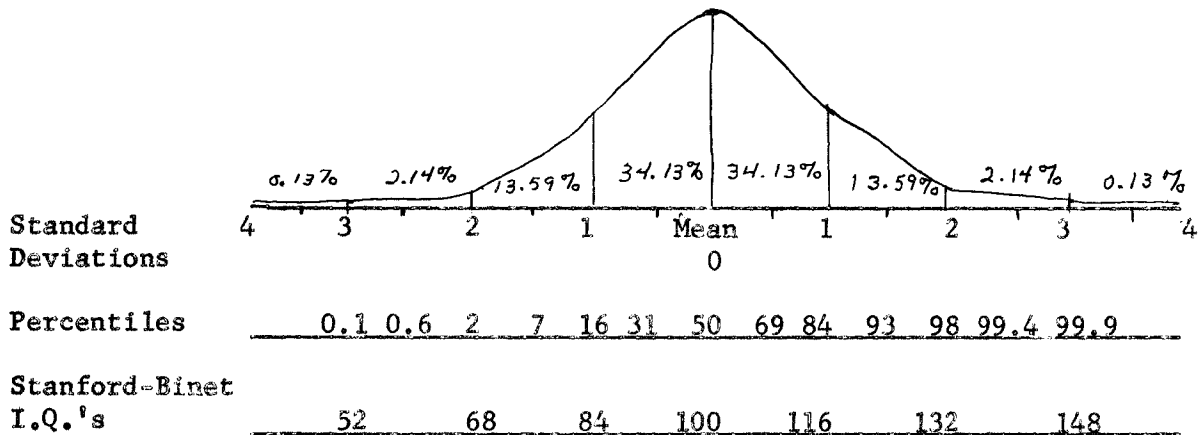
Informal Evaluation - Judgment of student progress which is based on subjective observation.

Intelligence Quotient - A measure of potential rate of growth up to sixteen years of age, expressed as the ratio of mental age to chronological age. The formula is: $I.Q. = \frac{M.A.}{C.A.} \times 100$. For ages over sixteen years, 192 months is used as the chronological age, on the assumption that, on

the average, mental maturity does not increase materially with further increases in chronological age.

Item Clarity - The quality of a test which enables the student to understand exactly what is required of him.

Normal Probability (Distribution) Curve - A derived curve based on the assumption that variations from the mean are by chance. It is bell-shaped, and adopted as true because of its repeated recurrence in the frequency distributions of sets of measurements of human characteristics in psychology and education.



Norms - Scores which the ordinary student of a certain age or grade would receive on a test. Norms are arrived at by giving the same test to a great many students. Common types are age, grade, and percentile.

Objectivity - A term used to describe a person's reaction, based upon facts alone, to something or someone.

Observation - The process by which information is received through the five senses.

Oral Test - A test in which the student is asked to express aloud his understandings of a subject, either alone or in a group.

Percentile - One of the 99 point scores that divide a ranked distribution into groups, each of which contains 1/100 of the scores. If a person obtains a percentile rank of 70, his standing is regarded as equaling or surpassing 70 per cent of the normative group on which the test was standardized.

Performance Test - A test in which the student displays his ability to perform the skill in which he has received instruction.

Rating Scale - (Also called "check list".) A device which is used to place the performance of a student into various levels of proficiency. Often a series of four or five numbers is used; for example, 0 - 1 - 2 - 3 - 4 where 4 is very good and 0 is unsatisfactory.

Raw Score - A term which refers to a test performance before it is converted to a percentage.

Reliability - The quality of a test which produces similar scores when readministered with no teaching in the intervening period of time and in exactly the same way.

Role - A term used to describe the function of a person or thing in relation to another person or thing.

Scoring - (Also called "grading") An activity which involves the interpretation of testing results in the light of certain definite criteria.

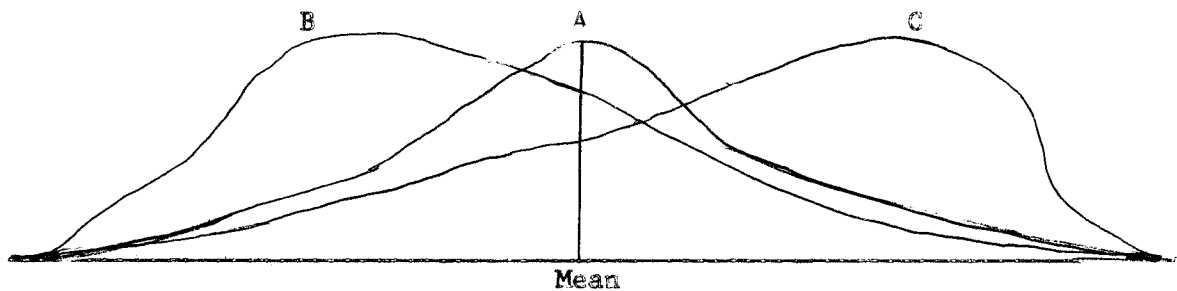
Self-Reports - Written accounts by the individual student dealing with certain topics.

Senior High School - An educational establishment in which instruction is normally given in grades ten, eleven, and twelve.

Short Answer Test - (Also called "objective test") A test in which factual information is required, usually in the form of choosing alternatives.

Specific types are (1) completion, (2) recall of a single fact, (3) recognition, (4) true - false, (5) multiple-choice, and (6) matching.

Skewed Curve - A term which refers to a statistical curve which falls either to the left or right of center because the group which it represents does not, for some reason, fall into the normal pattern. For example:



In the diagram, "A" represents the normal curve, while "B" and "C" represent curves that are skewed to the left and right respectively.

Skill - A term which denotes the efficient performance of mental or physical tasks.

Standard Deviation - A unit of comparative measurement based on the normal curve, and representing the distance from the mean which in one direction includes 34.13 per cent of the total cases.

Standard Error - A term which represents the relative amount of inaccuracy, depending upon the number of cases involved, in any statistic.

Standardized Test - A test developed by experts which has been given to so many students that it has been possible to determine reasonably accurately how well a typical student of a certain age or grade will do in it.

Subjectivity - A term which denotes a person's inner reaction, based on his own personality, feelings and experience, to something or someone.

Talent - A relatively high order of aptitude. It refers to an individual's susceptibility to an unusually high degree of training.

Validity - The quality of a test which measures what the test was designed to measure.

Work Sample - An excerpt from a course of studies which has been chosen for testing purposes. Time limitations require that it not be too long, and adequate representation requires that it not be too brief. A type of performance test.

Basic Assumptions

The study is based on the following assumptions:

1. It is possible to evaluate many areas of music education accurately and objectively.
2. Some areas defy objectivity; nevertheless, they can usually be evaluated by means of subjective techniques.
3. It is not necessary for evaluation in music education to be devoid of clearly defined standards of comparison.
4. Information found in books, professional periodicals, academic studies, and interviews with successful teachers is sufficient to give reasonable insights into the problems of evaluation.

Purpose of the Study

One of the purposes of this study is to clarify the question of evaluation in music education, particularly as it relates to the matter of grading student progress. It is disconcerting to see the matter constantly ignored or pushed into the background. At least an attempt should be made to explore the materials and techniques which are available;

to decide whether or not evaluation should be a part of high school music education; and, if so, to determine what forms it should take.

Secondly, it is hoped that this study will suggest to both experienced and prospective high school music teachers a workable system for the evaluation not only of student progress but also course objectives and teaching methods.

Thirdly, the study will acquaint teachers with the better types of evaluative materials and techniques which are available to them.

Method of Procedure

1. The literature relating to the problem was reviewed for ideas pertinent to the study.
2. The literature included books (psychological and educational), articles from professional periodicals, masters theses, and doctoral dissertations.
3. Investigation was made of available standardized tests by consulting the Educational Testing Service, 1947 Center St., Berkeley, California, Tests in Print,¹ and the Mental Measurements Yearbooks (from 1940 to 1965).²
4. Some ideas were gathered from formal and informal interviews with successful teachers.

¹Buros, Oscar Krisen (ed.) Tests in Print, Highland Park, New Jersey: The Gryphon Press 1961.

²Buros, Oscar Krisen (ed.) Mental Measurements Yearbooks, Highland Park, New Jersey: The Gryphon Press, 1940-1965.

5. Conclusions and recommendations have been derived from these sources.

Review of the Literature

M. B. Stanton³ of the Eastman School of Music in Rochester, New York, made a ten-year investigation of the Seashore Measures of Musical Talents. Over 2,000 entering students were tested, and the results were filed until four years later. It was found that where the Seashore tests had predicted success, teachers' evaluations confirmed this.

However, A. Richard Roby⁴ found that the correlation between the results of the Seashore tests and college theory grades was low, and even negative. On the other hand, he found correlations as high as .773 between the Aliferis Music Achievement Test and college grades. The Aliferis test, unlike the Seashore, uses musical multiple-choice items from which the student chooses the notation which he hears played on the piano.

P. R. Farnsworth⁵ found that validity coefficients in the Seashore tests, when compared to school grades in music, varied from -.08 to .45. Reliability was between .62 and .89.

³Super, Donald E., and Crites, John O. Appraising Vocational Fitness, Revised edition, New York: Harper and Brothers, 1962.

⁴Roby, A. Richard. "A Study in the Correlation of Music Theory Grades With the Seashore Measures of Musical Talents and the Aliferis Music Achievement Test," Journal of Research in Music Education, Vol. X, No. 2 (Fall 1962) 137-142.

⁵Farnsworth, P. R., "An Historical, Critical and Experimental Study of the Seashore-Kwalwasser Test Battery," Genetic Psychology Monograph (1931) 9: 291-389.

Hazel Stanton⁶ showed that I.Q. scores are at least as important in predicting success in music as the Seashore tests. In fact, when she used both the Seashore tests and the Iowa Test of Silent Reading, she was quite successful, as the following statistics show:

Group	Number	Percent of Graduated
Safe	125	60
Probable	143	42
Possible	195	33
Doubtful	73	23
Discouraged	29	17

A. M. Jordan⁷ found that in the practical and theoretical areas, intelligence tests were at least as valuable in predicting success in music education as the Seashore tests. This was not true in sight singing, ear training, or dictation.

J. C. Cooley⁸ conducted a study which showed evidence that high intelligence, high reading ability, and superior performance on the Seashore tests go with musicality. There was no evidence that personality traits are so related.

In the Minnesota Mechanical Abilities Project which was conducted by D. S. Paterson⁹ in 1930, industrial arts teachers rated shop products of junior high school boys. Reliability was found to be between .72 and .76.

⁶Stanton, Hazel. Measurement of Musical Talent: Studies in the Psychology of Music, Vol. II. New York: University of Iowa Press, 1935.

⁷Jordan, A. M. Measurement in Education, An Introduction. New York: Mc Graw-Hill, 1953, pp. 288-291.

⁸Cooley, John C. "A Study of the Relation Between Certain Mental and Personality Traits and Ratings of Musical Abilities," Journal of Research in Music Education, Vol. IX, Nor. 2 (Fall 1961) 108-117.

⁹Super, op.cit.

G. A. Forehead¹⁰ conducted a study in English Literature which involved understanding, interpretation, evaluation, and taste.

- a. Understanding. Multiple-choice tests of ten items each were used for testing recognition of literary devices and thematic material. (Significant results were noticed by the end of the year.)
- b. Interpretation. This involved an essay on any topic regarding the story in question. A reasonable degree of insight was expected.
- c. Evaluation Technique. A check list using different adjectives was employed; for example, valuable-worthless, fair-unfair, pleasant-unpleasant, positive-negative.
- d. Taste. This included (1) facts (about setting, plot, and character), (2) entertainment value (to amuse or create suspense), (3) skill of author (literary technique), and (4) theme (relevancy to human experience).

The test used reaction questions to literary excerpts. All alternatives were correct. Students were asked to pick the one they liked best. By the end of the year, a shifting away from "fact" alternatives was observed.

¹⁰Forehead, Garlie A. "Problems of Measuring Response to Literature," The Clearing House, Vol. 40, No. 6 (February, 1966) 369-375.

K. U. Gutsch¹¹ conducted a study in which mathematics was combined with sight reading in order to determine whether objective measurement of instrumental music achievement could be obtained. Three statistical implications essential to the study were:

1. Reliability in test and scoring.
2. Items were re-ordered and retested to increase reliability.
3. Validity in that the test differentiated between students with varying degrees of experience.

The results indicated that the test reflected both test and scoring reliability. Validity was also good. Results showed that student experience was the most influential factor governing performance. Age had relatively little effect upon ability to sight-read rhythms.

R. J. Colwell and Glenna Rundell¹² conducted a study to determine what effect the addition of (1) the ukulele and (2) the piano would have upon achievement in auditory-visual discrimination tests, harmonic conceptualization, better understanding of uses for factual knowledge, and attitude toward music in general. Three similar grade seven classes were used; one with ukuleles, one with pianos, and one with a vocal approach supplemented with bells and autoharps.

Various standardized tests were used in pretesting and post-testing, and the result was that the class using ukuleles developed a more favorable

¹¹Gutsch, Kenneth U. "Instrumental Music Performance: An Approach Toward Evaluation," The Journal of Educational Research, Vol. 59, No. 8 (April, 1966), 377-380.

¹²Colwell, Richard J., and Rundell, Glenna. "An Evaluation of Achievement in Auditory Discrimination Resulting from Specific Types of Musical Experiences Among Junior High School Students," Journal of Research in Music Education, Vol. XIII, No. 4 (Winter 1965) 239-245.

attitude towards music, However, no significant value was found in using ukulele or piano over the vocal approach. In fact, the keyboard and vocal approaches showed more lasting learning than the ukulele approach when tested a year later.

The "Hawthorne effect" was very evident in this study. That is, teachers were honored to have a part in the study, and worked harder to relate objectives to evaluation.

E. L. Rainbow¹³ conducted a study to discover the factors involved in musical aptitude. The following were checked: (1) pitch discrimination, (2) tonal memory, (3) rhythm, (4) musical memory, (5) academic intelligence, (6) school achievement, (7) sex, (8) chronological age, (9) musical achievement, (10) musical training, (11) home environment, (12) interest in music, (13) participation in music by relatives, and (14) socio-economic background. All but school achievement, sex, and participation in music by relatives were found to have a significant bearing on aptitude. Age had considerable influence.

J. H. Fluke¹⁴ conducted a study in which it was assumed that the quality and extent of the performer's awareness or perception into the inner content of music is a basic constituent of musicality. Accordingly, a test was constructed which measured basic rhythmic, melodic, and harmonic concepts by having students listen to excerpts of music which

¹³Rainbow, Edward L. "A Pilot Study to Investigate Constructs of Musical Aptitudes", Unpublished Ph.D. dissertation, State University of Iowa, 1963.

¹⁴Fluke, John Holman, "The Construction, Validation and Standardization of a Test in Music Perception for High School Performance Groups," Unpublished Ph.D. dissertation, Colorado State College, 1963.

illustrated each of these concepts, and answer thirty multiple-choice questions on each.

The test was called the Fluke Test in Music Perception. It was administered to 2,314 high school instrumental and choral students. The reliability coefficient for the test as a whole was found to be .78, which is satisfactory.

It was found that instrumental students were more perceptive, and hence more musical than choral students.

Colwell and Rundell¹⁵ constructed a test which consisted of thirty pairs of music excerpts from which the student was asked to choose the one which he considered to be best performed.

It was found that the instrumental and experienced student had better aesthetic judgment.

L. M. Hagen¹⁶ conducted a questionnaire survey in which choral teachers were asked to indicate the bearing on their grading of the following factors: attitude, cooperation, effort, enthusiasm, responsibility, attendance, promptness, talent, general musical knowledge, tests, sight-reading, memorization, private lessons, and outside activities. Teachers were also asked (1) whether or not they used a point system in grading, (2) whether or not students were informed as to what was expected of them for A, B, or C grades, (3) how their choral grades compared with

¹⁵Colwell and Rundell, op.cit.

¹⁶Hagen, Lawrence, M. A Survey of Choral Music Grading in High Schools of Five Hundred Students or More in the State of Washington, Unpublished masters thesis, University of Montana, 1962.

academic grades in the school, and (4) whether or not they were satisfied with their present grading system.

This survey indicates (1) that grading in choral music is very subjective, and is based more on extra-musical factors than skill, progress, proficiency, or musicianship; (2) that respondents who used the point system, where grading points are given or deleted according to a pre-scribed list of positive or negative activities, were generally satisfied; (3) that respondents who were dissatisfied usually had no system at all, and were not sure what to do about it; (4) that many respondents expressed the desire for more objectivity in evaluation; (5) that choral grades are generally higher than academic grades, principally because students who sing are often more intelligent; and (6) that those respondents who used the quartet system of grading praised it highly. (See Chapter V.)

Edwin Gordon¹⁷ conducted a study which showed no evidence that training improved scores on the Drake Musical Aptitude Test. This seems to suggest that musical aptitude, as measured by this test, is innate.

J. Hoffren¹⁸ conducted a study in which an attempt was made to test expressive phrasing in music. The following factors were considered: rubato, smoothness, articulation, phrasing, unity, continuity, dynamics, and dynamic and agogic accentuation. Although validity was not high, the test was moderately successful.

¹⁷Gordon, Edwin. "A Study to Determine the Effects of Training and Practice on Drake Musical Aptitude Test Scores", Journal of Research in Music Education, Vol. IX, No. 1 (Spring, 1961) 63-74.

¹⁸Hoffren, James. "The Construction and Validation of a Test of Expressive Phrasing in Music", Journal of Research in Music Education, Vol. XII, No. 2 (Summer, 1964) 159-164.

F. W. Pinkerton¹⁹ conducted a study in which methods of choosing instrumental students were investigated. It was found that commonly used criteria were (1) student and parent interest, (2) recommendations of classroom and general music teachers, (3) mental rating, (4) tests of musicality, (5) success in a pre-instrument class, (6) physical traits, (7) simple singing ability, (8) coordination, (9) scholastic standing, and (10) achievement test results.

"The most evident conclusion of this study is that there is little agreement among music psychologists, teachers, and supervisors as to methods of selecting instrumental music students in the public schools."²⁰

It was recommended that students be selected on the basis of not one, but several criteria.

¹⁹Pinkerton, Frank W. "Talent Tests and Their Application to the Public School Instrumental Music Program", Journal of Research in Music Education, Vol. XI, No. 1 (Spring, 1963) 75-79.

²⁰Ibid., p. 78.

Brief Historical Background to Evaluation in General

"Before 1900, teachers had very limited methods for determining how well children were succeeding."²¹ Completion and multiple-choice tests did not become common until about 1910. Many standardized aptitude and achievement tests, as well as intelligence tests, were developed during the 1920's, the "goldrush era" of standardized tests. Since 1940, educators have used other evaluative techniques such as anecdotal records, sociograms, rating scales, participation charts, case studies, cumulative records, personal interviews, check sheets, and report cards. One reason why so many evaluative techniques have been developed and used in music education is the conviction among many educators that musical talent is actually composed of many abilities rather than only one. "The rich assortment of skills found when we survey all varieties of musical experience suggests strongly that many subtalents, rather than a single all-pervading one, make up musical success in its broadest sense."²²

²¹Thomas, R. Murray. Judging Student Progress, New York: Longmans, Green and Co., 1954, p. 11.

²²Thomas, R. Murray. Individual Differences in the Classroom, New York: David McKay Co., Inc., 1965, p. 384.

CHAPTER II

The Nature of Evaluation in Music Education

Evaluation in music education is basically a matter of observation. The music teacher observes from day to day the reactions of his students to the listening or appreciation program, the progress of his instrumentalists, and the vocal quality produced by his choral classes. In fact, the good music teacher is the one who has learned to observe a great many different things, virtually at the same time. These observations must not only be made, but also recorded and converted into meaningful periodic reports of progress for the benefit of student, teacher, parent, and administrator.

What factors must be observed? The answer depends entirely upon the aims and objectives of the music teacher himself. Elliot W. Eisner says that evaluation is a "judgment of the adequacy of behavior as compared to a set of educational objectives."¹ This quotation implies that the teacher must (1) be certain about his objectives, (2) state them in terms of student behavior, not teacher behavior, and (3) state them clearly enough that it is possible to tell when they have been achieved. Occasionally it may be deemed wise to shift from original objectives. It may occur to the sensitive teacher, for instance, that a particular class of student would perhaps profit from a different approach to the course material. In such an event, evaluation would shift its emphasis accordingly.

¹Eisner, Elliot W., "Evaluating Children's Art," School Arts, 63 (September, 1963), p. 20.

Evaluation, however, is not simply observation. Although some evaluation can be made by simple observation, it is often erroneous or incomplete. More systematic methods are necessary. At the same time, the term "evaluation" is broader in meaning than the term "measurement", which implies the use of conventional tests and examinations. Measurement is the part of evaluation which concerns itself with subject-matter achievement or specific skills and abilities. Evaluation concerns itself with certain educational objectives and "the appraising of behavioral and personality changes which result from the educational program."²

What are some of the "educational objectives" which were referred to in the preceding paragraph? The following represent some examples: (1) In singing, the ability to use the voice to express beauty in song together with artistic interpretation. (2) In choral work, the ability to use one's vocal skill in conjunction with others. (3) In music appreciation or the listening program, to develop discrimination and taste, to develop sensitivity to design, balance, quality, and appropriateness, and to acquire a general knowledge of the development of Western music (including such factors as music history, style, great composers and their works, harmony, form, and orchestral instruments). (4) In instrumental music, the ability to express oneself, at least to some degree, on a musical instrument. (5) In creative music, some degree of originality in interpreting and composing music. (6) In musical scores,

²Leonhard, Charles, "Evaluation in Music Education," The Fifty-seventh Yearbook of the National Society for the Study of Education, LVII, Part 1, (1958), p. 310.

the ability to read music, ability to use musical notation to express musical meaning, and ability to work with certain phases of music theory such as accent, bars, phrases, scales, chords, staves, key signatures, and time signatures. (7) In more intangible phases of music education, the development of such factors as freedom of expression, rhythmic sense, and ability to function effectively in music groups or organizations. "Only when the teacher has clearly determined his goals can he select appropriate teaching methods to the exclusion of irrelevant material or teaching techniques."³

The teacher who has many objectives and many approaches will have many means of testing. It is the balance between these factors which produces superior results, and perhaps this delicate balance can be achieved only through the teaching experience of the individual teacher. "Since the purposes for testing and the objectives of instruction vary, it is readily apparent that the types of tests used in measuring achievement must also vary. There is no one best type of test."⁴

"Tests are not alternatives to observations. At best they represent no more than refined and systematized processes of observation."⁵ Normally, tests are based on the "work sample" principle. This sample must be truly representative, and large enough not to be effected by accidental factors. Actually, because of the existence of time limits, a test is

³Franklin, A. David, "Ends and Means in Music Education," Music Educators Journal, 53, No. 7 (March, 1967) p. 106.

⁴Green, John A., Teacher-Made Tests, New York: Harper and Row, 1963, p. 3.

⁵Ebal, Robert L. "The Social Consequences of Educational Testing," School and Society, 92 (November 14, 1964), p. 331.

"a sample of how well the student works with certain kinds of learning material."⁶ If a student works better, faster, and more accurately, he gets a higher score. No single test measures all of any ability. This is why many different samples of performance need to be taken. It is rather like letting down pipes into the ground at various locations in an effort to ascertain which direction an underground river is taking. Also, a test performance is susceptible to change. Young people are especially changeable. By working at a skill, for instance, the student can improve his score.

"Assigning marks or grades is one aspect of the appraisal of student progress."⁷ The value of this practice may be debatable, yet teachers generally are still required to do it. However, it must be done fairly, and be based upon all objectives, not merely one or two. It is extremely important to point out in this connection that what these grades represent must be made perfectly clear. When the student and parent receive a music grade, are they aware that it represents a comparison between the student's performance and that of the rest of the class, or between his performance and a standardized norm, or between his present performance and past performances?

In order to see how the student is actually progressing, some comparison must be made. Often a student's performance is compared with

⁶Dobbin, John E., "Still Testing, Testing, Testing," The P.T.A. Magazine, 60 (January, 1966), p. 5.

⁷Leonard, Charles, "Evaluation in Music Education," The Fifty-seventh Yearbook of the National Society for the Study of Education, LVII, Part I (1958), p. 312.

that of the rest of the group to which he belongs. If the standard of the group with which his performance is being compared is large enough, it is referred to as a "norm". Most "standardized" tests are published complete with norms for the teacher's consideration. In teacher-made tests, however, the teacher himself must develop his own norms, if he desires to use them, by giving his test to a great many students.

Actually, in order to determine real progress, the student's past and present performances must be compared. "If we want to know if a child has gotten taller it is fruitless to find out if he is above average in height."⁸ Using an individual comparative base of this kind also serves to improve instruction because it provides opportunities to detect weaknesses in performance. In cases where the student is below the norm, we may get away with blaming the student; but where no growth occurs from one week to the next, we are forced to look very carefully at our methods, materials, and objectives. This is not to suggest, however, that the student share no responsibility at all for lack of growth. Yet we must be prepared to do something to try and strengthen his weaker areas. In addition, the sooner a past basis can be decided upon for each student, the better. Where there is a record of achievement from pre-senior high grades, evaluation is much easier and more accurate.

Students should be encouraged to evaluate themselves in the light of past performances. Much encouragement can result from this approach.

⁸Eisner, loc.cit., p. 21.

Students should also be encouraged to write about their individual attitudes and interests in music. From such information the teacher may discover the areas of study upon which to concentrate and how to approach them. He may also discover significant changes as the school year progresses. Since it is generally agreed that the best teaching springs from student interest, the student who takes the time and trouble to evaluate himself periodically, is bound to make reasonable progress.

Evaluation in music education does not differ in any way from evaluation in other areas, and is essential in every classroom situation. In fact, since the music program involves a good deal of group activity, there is excellent opportunity for group evaluation of activities. Music "provides the circumstances for the most effective evaluation while work is in progress and by those immediately involved."⁹ While participating in group activities, the student is able to compare his performance not only with that of other students around him, but also with his own past performances. "Music activities call for the immediate application of principles. The development of skills takes place in actual performance and the development of knowledge and attitudes is an integral part of this performance."¹⁰

At the conclusion of this chapter it must be clearly understood that evaluation in music is intimately associated with instructional

⁹National Association of Secondary-School Principals Bulletin, XLII, No. 245 (March, 1959), p. 43.

¹⁰IBID.

objectives and methods. If, for example, the teacher is satisfied with instilling into his students mere factual information about their listening program, he will also be satisfied to evaluate results with a true-false or multiple-choice test. The teacher, on the other hand, who is concerned with helping his students understand some of the more intangible aspects of their listening program, such as atmosphere, style, or taste, finds that evaluation becomes much more of a problem. He finds that he must turn to more "subjective" instruments of testing, which at best, are vulnerable to considerable inaccuracy depending upon the personal opinions, attitudes, and background of the examiner. In the following chapters I shall try to point out the advantages and disadvantages of both objective and subjective evaluation techniques, and the respective values of each in various situations.

Purposes of Evaluation in Music Education

Evaluation is a means whereby the music teacher can measure the merits of his objectives. Objectives which cannot be measured are worthless; concepts which the teacher wishes to present to his students must be crystal clear in his own mind first of all if he expects to have worthwhile results. If these concepts are clear enough to be evaluated, they are likely to be clear enough to be understood by the student. "Evaluation provides the only avenue for determining the extent to which the program is caring for the overall musical development of all students."¹¹

Evaluation is also useful in helping the teacher to appraise his methods of instruction. If the results of evaluation in any given area at any particular time tend to display a rather discouraging rate of student growth, the teacher is well justified in looking quite critically at the manner in which he is approaching the material. Given that the objectives which he has in mind seem to be reasonable, or have seemed satisfactory in the past, it may be that for this particular student or group of students his point of view or point of departure, for example, is not right. In a case like this there is certainly nothing wrong with starting all over again and evaluating the results again, perhaps in a different way. Once the results reach more encouraging levels, the teacher can assume that his methods now are more in line with the present situation.

¹¹Leonhard, loc.cit., p. 314.

Evaluation should increase motivation. It should encourage the student to work harder, and better his present performance not only through comparison with other students or past performances, but also through comparison with a concept of a realistic level of aspiration. This concept might be developed in a student through the listening program, attendance at live concerts, or through the study of more advanced musical scores. Motivation also depends upon the degree of understanding and acceptance of the teacher's objectives. Effective evaluation may serve not only to clarify concepts and awaken interest, but may also help to make the student aware of his progress towards these objectives. In addition, the type of evaluation for which the student prepares determines in great measure the nature of his work. If, for example, the teacher wants the student to learn concepts which can be evaluated by means of a true-false testing instrument, that is one matter; but if, on the other hand, the desired concepts can be tested only by other, more subjective means, that is quite another. The point is that the objectives differ in each case, and so do both the nature of the learning and the means of evaluation.

In addition to those purposes of evaluation which have just been discussed, there remains that of student guidance. The teacher should make use of some form of evaluation, not the least of which should be standardized tests especially designed for this purpose, to help him decide which students should (1) take a high school music course at all, (2) take a course in instrumental music, (3) take a course in choral music, (4) take general music, (5) consider post-high school music

education, (6) consider music as a career. There are standardized aptitude tests which would be useful in conducting such guidance procedures. They are discussed in some detail in Chapter IV, and are listed in Appendix I.

Limitations of Evaluation in Music Education

In general, the objectives of music education break down into three basic areas: knowledge, attitudes, and skills. The first deals with objective facts, principles, and concepts. Of these, facts are the easiest to test: objective measuring instruments such as multiple-choice, completion, true-false, and matching tests are accurate and relatively easy to grade. Principles and concepts, however, because they are more difficult to evaluate accurately, must be measured by means of essay questions or problems requiring application of knowledge to specific situations. The second area, attitudes, involves feelings and emotions, are consequently difficult to teach, and are so difficult to measure that most teachers hesitate even to try. The third area, skills, involves both neuro-muscular learning and facility in the application of factual knowledge. These are also difficult to evaluate.

A single test is severely limited, almost to the point of meaninglessness. ". . . There is little hope of proving anything in education with single measures."¹² The teacher must test his students many times throughout the year, preferably with the aid of a variety of testing instruments, both objective and subjective, standardized and teacher made. The more often tests are given, the smaller the standard error, and the more dependable, meaningful, and accurate the evaluation.

The results of testing can be misleading unless the following aspects of test construction are taken into account: (1) Mechanical Aspects. The

¹²Educational Testing Service. Short Cut Statistics for Teacher-Made Tests: Princeton: Educational Testing Service, 1960, p. 20.

student should be aware immediately what he is required to do and how he is expected to do it. In order to ensure this attribute, similar items should be grouped together, directions should be clearly stated, and the organization of the entire test should be obvious. (2) Validity. The test should set out to evaluate only the objectives of the course, and give the most weight to the more important issues. (3) Reliability. The test should be so constructed that the results would not vary significantly in subsequent administrations to the same student (assuming no additional learning or practice occurs), or in subsequent gradings by the same teacher. (4) Appropriateness. Each item should be suitable to its objective, as well as to the age and grade level of the student for whom it is intended. (5) Clarity. The student who knows the material upon which the test is based should be able to understand the questions. (6) Discrimination. Test items should not, in any way, favor the student who has not really met the objectives of the course. When preparing a test, the teacher must constantly be on the lookout for grammatical or structural clues, opportunities for guessing, hidden answers elsewhere in the test, or obvious items.

So far, we have said little about grading. Yet the manner in which a test is graded can greatly influence the dependability of the results. Some principles to be considered in this connection are: (1) Grades should reflect, as far as possible, the actual achievement of each student, not merely innate ability or attractive personality. (2) Grades should not be used for disciplinary purposes. (3) Students should be acquainted with the grading system and with each grade assigned to them throughout the year. (4) There should be staff consensus on grading policy.

(5) Grades should not vary substantially from normal probability curve percentages unless the class is unusually bright or slow, in which case the curve would be skewed.

Grading is particularly difficult in essay-type test items. For instance, whether or not the examiner considers in his grading such extra-musical factors as spelling, sentence structure, punctuation, grammar, appearance, and paragraphing, has considerable influence upon the validity and reliability of the item. Although these things have nothing to do with music education per se, few would discount their importance in educating the student for life. Moreover, if the music teacher considers the student's capacity for English expression to be a part of his objectives, he is certainly justified in considering this in his evaluation. (See page 40 for specific suggestions regarding essay evaluation.)

In order for grading to be accurate, it must be based upon definite criteria. Three possible standards are: (1) the individual standard, in which the individual student's ability is taken into account, (2) the fixed standard, which is based upon the mastery of subject matter (In this case, the needs, interests, and abilities of each student are ignored, and the standard which is considered to be either satisfactory or unsatisfactory may vary considerably from one teacher to another. In one part of the country the pass mark may be seventy per cent, while in another it may be forty per cent.), and (3) the group standard or percentile, in which comparisons are made with the whole group according to the normal probability curve or standardized norms.

Finally, one must realize that the distinction between objectivity and subjectivity in music evaluation exists only at the grading stage. Both types of evaluation are subjective at the setting stage. The kinds of questions asked, whether they be true or false or essay type, are very much a matter of personal judgment and preference on the part of the examiner. Consequently, grading should be as objective as possible in order to avoid serious loss of accuracy and reliability. This objectivity in grading can be ensured through the use of aids such as standardized norms, the normal probability curve, and other clearly understood standards of comparison such as those referred to in the previous paragraph.

CHAPTER III

Two General Methods of Evaluation in Music

Formal

"Formal evaluation implies the use of an evaluative tool that is standardized and meets acceptable levels of reliability and validity."¹ Although some music educators would argue that much in music is so subjective and intangible that it cannot be measured in the same formal manner as other academic school subjects, Daniel Bonade, former principal clarinetist with the National Broadcasting Company Symphony Orchestra, believes that a subjective musical element such as phrasing can be taught and recognized objectively just as well as technique.² (See page 18 of "Review of the Literature".) If this is true, the classroom teacher should be able, at least to some degree, to use formal evaluative instruments to compare student performance with prescribed criteria and group norms in order to arrive at definite scores.

In the past, formal testing in music education has been avoided for the following reasons: (1) Desire to maintain a "fun" atmosphere. (Yet students in other subjects often enjoy their studies in spite of formal testing.) (2) Lack of generally accepted goals. (As a result, we are not sure what to test for.) (3) A failure of well-known standardized music tests to prove themselves to be either valid or reliable. (However,

¹Colwell, Richard, J. "Evaluation: Its Use and Significance," Music Educator's Journal, 49 (February, 1963), p. 47.

²Ibid., p. 49.

this is true mainly in the field of musical aptitude which has not yet been adequately defined. Failure in this area does not necessarily imply failure in other areas). (4) Teachers are often unaware of formal measures available to them. (5) The tendency towards a more subjective, aesthetic personality make up of most music teachers.

Specific information relative to formal, standardized tests in music is to be found in Chapter IV and Appendix I.

Informal

Many music educators feel that the study of music offers such a wide range of challenging objectives, that some can only be evaluated subjectively and informally, inadequate and inaccurate as this may be. Examples of informal evaluation are (1) casual observation, which is really a general impression, and could be quite wrong. (This method may involve, for example, the formation of impressions regarding care of an instrument, posture, enthusiasm, degree of cooperation, attitude, individual practice, general participation, and ability to work with others.) (2) anecdotal records, or brief happenings recorded by the teacher, and (3) time-sampling, which is a recording of student activities during a particular part of the day over a defined period of time. (In order for this device to have any significance, there must be many samplings.)

A rather important means of informal evaluation is the Rating Scale, also known as the Check List. This is a device by means of which the performance of a student can be placed at various levels of proficiency. Often a series of four or five numbers is used in which, for example, 5 represents excellent, 0 represents unsatisfactory, and the digits in between

represent the various intermediate gradations. This device has advantages in (1) diagnosing student strengths and weaknesses, (2) helping students to evaluate their own progress, and (3) helping to report progress in a meaningful way to parents and administrators.

Still another form of informal evaluation involves the use of the Cumulative Record. This is a collection of information about a student which covers his entire public school career. It is usually kept in the school office or guidance department where it can be contributed to and checked by all teachers. Normally, it has a face or summary sheet at the beginning which gives data such as psychological test scores (including aptitude, preference, and intelligence quotient), health reports, attendance, academic grades, schools attended, age, family and home background, and social adjustment. The rest of the folder has important anecdotes, teacher's comments, samples of work, results of interviews, and other information which could be of interest to the educator. Some systems use cards rather than folders.

"Informal evaluation is a necessary part of the total evaluation picture, which must include also formal evaluation, characterized by objectivity and systematic controls."³ Subsequent chapters shall discuss the issue of the most effective balance and proportion between formal and informal evaluation; in other words, between objectivity and subjectivity.

³IBIS, p. 46.

CHAPTER IV

Two Specific Types of Tests and Testing Procedures

Objective

The term "objective test" includes both aptitude and achievement standardized music tests. A standardized test is one which has been given to a sufficiently large number of students that it has been possible to determine norms. It has also been carefully checked to ensure validity and reliability. Tests which are published with no information as to their validity or reliability are not recommended. Reputable tests are constructed by experts, and are printed and distributed by test agencies, book publishers, and universities. Appendix I consists of a list of recommended standardized aptitude and achievement tests.

"All aptitude tests are, to some degree, achievement tests. An aptitude test may be distinguished from an achievement test only to the extent that the generalized function of aptitude is relatively maximized and specifically taught course-content material is relatively minimized."¹ For all practical purposes, however, aptitude tests are designed to measure the student's innate musical talent and to predict his future success in music education. On the other hand, achievement tests are designed to measure what the student has learned, in order to give the teacher some point of departure for his educational objectives. Both types are helpful, moreover, in grouping students for purposes of instruction. The Wing Standardized Tests of Musical Intelligence are designed to measure musicality and musical sensitivity, and are especially valuable in helping to select students for special instruction.

¹Gordon, Edwin, "The Musical Aptitude Profile," Music Educators Journal, 55, No. 6 (February, 1967), p. 52.

One of the most widely known and used standardized tests is the Seashore Measures of Musical Talents. It consists of one series of tests for unselected groups in general surveys, and a second series for musicians and prospective or actual music students. Interestingly enough, Hazel Stanton conducted a series of studies which appear to demonstrate that intelligence quotient scores are at least as important in predicting success in the study of music as the Seashore tests.² The teacher who, for one reason or another, does not make use of the Seashore or other aptitude tests, might find it of value to try to attract into his program as many students with high intelligence quotients as possible. "General ability is a sign of probable superiority in most types of achievement."³ Also, checking by the teacher on student success in other activities, particularly artistic ones, is well worthwhile. "Published tests are convenient, but they are not essential. The teacher who understands what might cause learning difficulty can make a sound diagnosis by observing a pupil individually."⁴

The use of aptitude or talent tests depends upon whether or not one's educational philosophy allows that all children should be offered a musical education; or that instruction, particularly at the high school and post-high school levels, should be limited to the promising few. One should bear in mind, in any event, that other factors besides

²Stanton, Hazel, Measurement of Musical Talent: Studies in the Psychology of Music, Vol. II, New York: University of Iowa Press, 1935.

³Granbach, Lee J. Educational Psychology, New York: Harcourt, Brace and Co., 1954, p. 200.

⁴Ibid., p. 176.

capacity may determine success in musical activities; for instance, motivation, level of initial and sustained interest, and the degree of aspiration. "Few individuals acquire sufficient mastery of the voice or a musical instrument to make music a profession. All students and adults, however, can acquire a taste for good music and can become intelligent listeners."⁵

The term "objective test" also includes some teacher-made tests, principally of the true-false, multiple-choice, completion, recall of a single fact, and matching variety. These tests are objective in the sense that they can be graded relatively accurately, but at the same time they are limited in respect to the types of educational concepts which they can effectively evaluate. Of these tests, multiple-choice "is the testing method most uniform, reliable, consistent, and impartial that we have at present."⁶ An interesting variety of multiple-choice is a test in which the possible endings include variables from best to worst, and the student is required not only to choose the best ending, but also to rate the endings in order of merit. Although it takes more time and trouble to set, it eliminates guessing, it can be used over and over again with only minor adjustments, and because it is relatively easy to grade, it is valuable when the teacher is working with large classes. By way of example, this type of test is an excellent means of measuring student's knowledge of clarinet or violin fingerings. Other

⁵Adams, Georgia Sachs; Torgerson, Theodore L; and Wood, Ernest R. Measurement and Evaluation for the Secondary-School Teacher, New York: The Dryden Press, 1957, p. 406.

⁶DePue, Pelmer, "Multiple Choice and the Either-Or-Fallacy," School and Society, 93 (March 6, 1965), p. 156.

examples of objective testing are the following: (1) The student looks at a sheet of music, the examiner plays it, and the student marks where the examiner stops. (2) The student looks at a sheet of music and indicates on it where the examiner has played wrong notes, wrong rhythms, or wrong dynamics. (Notice that both (1) and (2) represent a means by which the teacher is able to evaluate sight reading ability in an entire group of students.) (3) The student listens to polyphonic music and indicates what the lower voice is doing. (4) The teacher plays a piece of music either on the piano or record player, and asks the students about such matters as form, time signature, instrumentation, mode, name of selection, and composer.

Subjective

On the other hand, there are many valuable concepts in music education which can be evaluated only through the use of "subjective" testing procedures. Perhaps the most widely used and well-known form of the subjective test is the "essay type". Although it tends to be less reliable than objective testing techniques, principally because of poor grading procedures, the essay-type test is strong in the areas where objective tests are weak; that is, where the evaluative situation does not lend itself to a choice between two alternatives. The essay question gives the student the opportunity to explore many avenues of the subject, and by so doing exhibit to the teacher his depth of understanding and scope of knowledge. It also allows him considerable freedom of expression and creativity. "Essay tests are especially helpful when the teacher

wants to observe how the student organizes his thoughts or how he arrives at conclusions."⁷

There are different types of essay questions ranging from the very brief to the very lengthy. The "restricted-response" or short-essay question has high reliability if there are a good number of them in a test, and they can be graded almost as accurately as objective questions. The longer essay question is called "extended-response" or discussion question, and although it allows the student a great deal of freedom of expression, it is difficult to grade reliably. Definition and outline questions are also considered to be essay type.

The matter of grading essay questions is of the utmost importance. In fact, the teacher must not only prepare the questions, but also the answers and the grading scheme. Two good methods of grading are (1) point-score, and (2) sorting. In the first method, the teacher decides on the number of points for each question or part question. He then takes one question and reads all the answers to that question. Finally, he assigns a grade to each test on the basis of established norms, percentages, or the normal probability curve. In the second method, all the tests are read quickly and placed in a predetermined number of piles representing different letter grades. Finally, the papers are reread to check for accuracy. Although both of these methods can be used reliably, the first is preferable.

⁷Cronbach, loc.cit., p. 505.

Another type of subjective evaluation is the oral test. This technique is rather time consuming, but it gives an extensive picture of the depth and scope of the student's knowledge. It is also useful as an instructional device, and helpful to students who, for some reason, have difficulty with written examinations. On the other hand, some students are too nervous or self-conscious for this type of evaluation.

As in the case of the essay test, proper planning is required. A pre-planned check list or rating scale should be used, such as the one which is illustrated in the next chapter in the discussion of instrumental and vocal evaluation. In spite of careful planning, however, oral testing lacks accuracy and reliability, and should be used only to measure factors which cannot be evaluated as effectively in any other way.

A third type of subjective test is the performance test. Like the oral test, it can be used as a learning and guiding device, and may be helpful to students with verbal deficiencies. Its advantage is that it can measure skills and abilities which it is not possible to measure efficiently in any other manner.

Performance tests are capable of measuring two factors; i.e., (1) skill or technique, and (2) the product or result. The teacher may consider three approaches to this kind of evaluation: (1) Identification. This approach stresses the product or accompanying factors of the skills. For example, the student may be asked to identify the parts of his particular instrument. The disadvantage of this approach, considered in isolation, is that the student may do very well in this, yet be relatively poor in the technical aspect of performance. (2) Simulated Conditions.

This approach emphasizes actual procedures and conditions. For example, how would the student use his musical skills and knowledge in a practical way through actual participation in musical activities? (3) Work Sample. Here the emphasis is on skill or technique. The student is required to perform a piece of music.

Once again, proper setting and scoring is of extreme importance. Complete analysis must be made of the desired performance first of all. Then, once the teacher has decided, in the light of course objectives, which musical factors he expects to see or hear in the performance test, he may make up a check list such as that illustrated in the next chapter or a rating scale using numbers from one to five to represent various levels of general proficiency.

Finally, other subjective methods of evaluation are: (1) asking students to discuss the differences in style and structure between a piece of music which they have just played or heard, and one with which they are already familiar, (2) asking students to identify the style, form, or mood of an unfamiliar piece of music to which they have just listened, (3) asking students to comment upon the form, ensemble, realism, vocal technique, source, and musical support of the drama in an excerpt from an opera or operetta, (4) asking students to tell which two of four pieces or excerpts are stylistically similar, and (5) asking students to prepare original compositions of their own in various styles.

CHAPTER V

Suggested Methods of Evaluation and Specific Types of Tests and Testing Procedures in Three Areas of High School Music Education

General Music

As in the case of most musical knowledge and activity, there is much overlapping among various areas of study. Although one ordinarily thinks of music appreciation in the general music setting, it is also very much a part of the choral or instrumental class. Music appreciation may be defined as "the apprehension and enjoyment of the aesthetic import of music."¹ "Appreciation of the expressive import of music is revealed in the quality of performance, in the ability to make valid value judgments of performance and composition, and, to an extent, by the degree of absorption a person exhibits during musical experience."² The last part of this quotation might be evaluated by means of simple observation. The other two factors might be measured by means of subjective testing procedures such as those discussed in Chapter IV.

However, it seems to be practically impossible to measure adequately the degree of music appreciation as such. One can measure ingredients of it, such as listening habits and preference, by means of informal methods such as casual observation, interviews, and questionnaires on leisure activities. It is helpful for the teacher to know such things as (1) how much the student sings or plays outside of school, (2) the nature of the radio programs and records he selects, (3) his general attitude

¹Leonhard, loc.cit., p. 330.

²Ibid., p. 331.

towards good music, (4) his intentions relative to continuing to be active in musical activities after leaving high school, and (5) the extent to which music serves as an emotional outlet.

Another factor which helps to evaluate the level of music appreciation is the student's ability to remember a melodic line or rhythmic pattern accurately. This ability is obviously associated with intelligent listening. It can be measured by means of the Drake Test of Musical Memory. It can also be measured by teacher-made tests in which the teacher plays short melodies or rhythm patterns, and asks students to sing or clap each one from memory. Such testing can be graded quite objectively, although it is easier to administer individually than collectively, and it requires that the examiner be able to play the piano or some other musical instrument.

In advanced high school classes, the teacher may consider it of value to have students learn to recognize chords and modes. Students might even find elementary forms of melodic and harmonic dictation an intriguing challenge. Naturally, such activity would imply considerable background in music theory. Evaluation of such skills would be an integral part of the learning process, and would also serve as an indicator of the student's level of appreciation.

Some educators feel that the objectives in listening skills constitute enjoyment and desire to listen. Others believe that it should also include " . . . ability to discriminate in such matters as melody, rhythm, and tempo, and to apprehend large tonal patterns. . ."³ The

³Ibid., p. 327.

Oregon Tests for Musical Discrimination measures the student's ability to discriminate between a masterpiece and its mutilated versions. Originally they were not standardized, but have since been revised at Indiana University.

Standardized aptitude tests are useful in general music primarily at the beginning of a term in order to (1) ascertain the capacity of each student to profit from a music course, (2) help group students into homogeneous classes, and (3) help to determine realistic course objectives for each class. Standardized achievement tests are useful at the beginning of a year for much the same reasons, and also at various times during the year to help measure student progress, the merits of course objectives, and the effectiveness of teaching techniques. All of the teacher-made objective testing devices which are discussed in Chapter IV, except of course, performance tests, are applicable wherever course objectives warrant.

Choral Music

In a choral music class, emphasis tends to be upon the practical aspects of vocal group activities; that is, upon actually learning how to sing well in a group, learning a repertoire of good choral music, and performing in public at various occasions during the year. Evaluation may take various forms, both formal and informal, objective and subjective, depending upon specific objectives, capacity of the class, and the personality of the teacher. Standardized aptitude and achievement tests may be used for the same purposes as those outlined in the section on general music. In addition, performance tests such as those described in Chapter IV are useful.

Music reading is generally considered to be a rather important aspect of a choral-music course, since one's effectiveness in any choral organization is determined largely by one's ability to read quickly and accurately. Music reading is, in fact, an audio-visual skill which involves the ability to relate what one sees on the printed page to musical sound. Ordinarily, the student's ability to read music must be evaluated orally, on an individual basis. However, if it is considered in reverse; that is, from the musical sound to the printed page, evaluation can be effectively performed on an entire group. Two standardized tests which could be used to measure reading ability are the Knuth Achievement Tests in Music and the Farnum Music Notation Test. In addition, the teacher can achieve comparable results through devices similar to the following: Copies of musical excerpts are mimeographed, given to the students to examine, then played on the piano. The students may be asked, for instance, to tell whether or not the excerpt has been played correctly, and, if not, the nature of the errors made. The advantage of this type of test is that there are a variety of ways in which it can be adapted to the particular needs of the class.

In group singing, the teacher may make use of informal evaluation by observing and recording (perhaps by means of check lists, anecdotal records, or time-sampling) such factors as (1) participation and co-operation, (2) enjoyment, (3) attention (4) posture, (5) correct time, (6) correct pitch, (7) enunciation, (8) good vocal production, and (9) memorization. Although these observations can best be made on an individual basis, this presents problems such as: what to do with the rest of the class, what to do about the self-conscious or nervous student, and how

to cope with the excessive time factor. On the other hand, it is difficult to arrive at a fair evaluation of a student's progress when there are many voices singing with him. Many teachers find the "quartet method" useful as a compromise between these two extremes. This device involves the choosing of four students, each of whom sing a different voice part (soprano, alto, tenor, or bass), and having them sing a choral selection. This selection of course, must be long enough to allow the teacher time to concentrate not only upon the ensemble, but also upon each voice. A check list such as the one below might be used to record reactions in simple percentages or letter grades. Incidentally, a eight by five-inch card could be made up for each student at the beginning of the year,

Name: Telephone:
Class: Gown:

Reporting Period	1	2	3	4
Sight Reading				
Technique				
Quality				
Blend				
Projection				
Diction				
Accuracy				
Musicianship				
Style				
Attitude				

consisting of a check list on one side and a register on the other for the purpose of recording student attendance at extra choral rehearsals.

Rehearsal and Performance Attendance

	First Week	Second Week	etc.
	M T W T F	M T W T F	
Sept.			
Oct.			
Nov.			
etc.			

There are standardized singing tests available, but all lack objective methods of scoring. The things which the musician or intelligent listener look for in good choral singing are subjective by nature, and must be evaluated with that in mind. For example, no objective method has been found to measure beauty of tone. "The artist, be he writer, painter, or musician, takes issue with this view of scientific infallibility, knowing that subjective values which cannot be measured or pinpointed are more enduring and of more personal importance than the objective, tangible values."⁴

Instrumental Music

Most of the matters discussed in the section on choral music apply here. Two possible differences, however, might be (1) that the instrumentalist is working with an instrument which is not as personal as the voice; hence, he is usually not as sensitive about performing alone as the vocal student, and (2) since it is relatively easy to keep instrumentalists busy in their practice rooms while the teacher is conducting individual tests, it is recommended that this be done at least once a year. The

⁴Colwell, Richard, J. "Evaluation: Its Use and Significance," Music Educator's Journal, 49 (February, 1963), p. 45.

following check list might be helpful in the administering of individual performance tests. As in the case of choral testing, a card could be made up for each band or orchestra student, the back of which would consist of an attendance record.

A more objective means of instrumental evaluation is found in the Watkins-Farnum Performance Scale for Winds and Strings. In this test, every error is evaluated, and an exact scoring system is given to help increase objectivity. It even includes a section for the evaluation of percussion. Since good reading ability presupposes good technical mastery,

Name:	Rental:
Class:	Telephone:
Instrument:	Uniform:

<u>Reporting Period</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>
Sight Reading				
Technique				
Quality				
Projection				
Accuracy				
Intonation				
Style				
Musicianship				
Attitude				

this test measures combined technical and reading skills to a greater degree than practical performance.

Conclusions

1. In the light of the variety of evaluative materials and procedures which are available to the high school music teacher, and the benefits which accompany their careful use, evaluation in a music program should not be minimized or ignored.
2. In order for evaluation to be accurate and dependable, it should be continuous and of considerable variety.
3. Some music values cannot be evaluated efficiently by means of objective measuring instruments.
4. Although objective evaluation is more accurate than subjective evaluation, some musical and educational values can be measured through subjective means only.
5. The proportion of objective to subjective evaluation in any testing program depends upon the educational objectives of the teacher.
6. There are undoubtedly some music values which are so personal that they cannot really be taught, or evaluated; yet, on the whole, the teacher should be prepared at least to attempt an evaluation of anything he teaches.
7. Intelligence quotients have been found by investigators to be of considerable value in predicting success in music education. (See "Review of the Literature".)

Recommendations for Further Study

1. An investigation of the actual evaluative procedures of successful high school music teachers.
2. An investigation into the relationship between evaluation in music and evaluation in the other arts.
3. A historical survey of evaluation in music education.
4. A futuristic study of evaluation in music education.
5. A controlled study in which a comparison in student progress and general results is made between a music class where a variety of evaluative techniques are frequently employed, and a class where evaluation is ignored.

BIBLIOGRAPHY

Articles

- Boutwell, William D., "What's Happening in Education?", The P.T.A. Magazine, 60 (April, 1966), 8-9.
- Clifford, Timothy F., "Skew That Curve," Music Educators Journal, 52 (November-December, 1965), 87-90.
- Coleman, Henry, "Examinations and Festivals," Music in Education, 26 (May 15, 1962), p. 72.
- Colwell, Richard J. "Evaluation: It's Use and Significance," Music Educator's Journal, 49 (February, 1963), 45-49.
- Colwell, Richard. "An Investigation of Musical Achievement Among Vocal Students, Vocal-Instrumental Students, and Instrumental Students." Journal of Research in Music Education, Vol. XI, No. 2 (Fall, 1963), 123-130.
- Colwell, Richard J., and Rundell, Glenna. "An Evaluation of Achievement in Auditory-Visual Discrimination Resulting from Specific Types of Musical Experiences Among Junior High School Students," Journal of Research in Music Education, XIII, No. 4 (Winter, 1965), 239-245.
- Cooley, John C. "A Study of the Relation Between Certain Mental and Personality Traits and Ratings of Musical Abilities," Journal of Research in Music Education, Vol. IX, No. 2 (Fall, 1961), 108-117.
- DePue, Palmer, "Multiple Choice and The Either-Or-Fallacy," School and Society, 93 (March 6, 1965), 154-156.
- Dobbin, John E., "Still Testing, Testing, Testing," The P.T.A. Magazine, 60 (January, 1966), 4-6.
- Ebel, Robert L. "The Social Consequences of Educational Testing," School and Society, 92 (November 14, 1964), 331-334.
- Eisner, Elliot W. "Evaluating Children's Art," School Arts 63 (September, 1963), 20-22.
- Forehand, Carlisle A. "Problems of Measuring Response to Literature," The Clearing House 40, No. 6 (February, 1966), 369-375.
- Franklin, A. David, "Ends and Means in Music Education," Music Educators Journal, 53, no. 7 (March, 1967), 103-106.
- Gordon, Edwin, "The Musical Aptitude Profile," Music Educator's Journal, 53, No. 6 (February, 1967), 52-54.

- Gordon, Edwin, "A Study to Determine the Effects of Training and Practice on Drake Musical Aptitude Test Scores," Journal of Research in Music Education, Vol. IX, No. 1 (Spring, 1961), 64-74.
- Gutsch, Kenneth W. "Instrumental Music Performance: One Approach Toward Evaluation," The Journal of Educational Research, 59, No. 8, (April, 1966), 377-380.
- Higgins, Martin J. and Merwin, Jack, C. "Assessing the Progress of Education in Music," Music Educators Journal, 53, No. 8 (April, 1967), 52-54, 131-132.
- Hoffren, James. "The Construction and Validation of a Test of Expressive Phrasing in Music," Journal of Research in Music Education, Vol. XII, No. 2, (Summer, 1964), 159-164.
- Leonhard, Charles, "Evaluation in Music Education," The Fifty-seventh Yearbook of the National Society for the Study of Education, LVII, Part 1 (1958), 310-338.
- National Association of Secondary School Principals Bulletin, XLVIII, No. 245, (March, 1959), 43-45.
- Pinkerton, Frank W. "Talent Tests and Their Application to the Public School Instrumental Music Program," Journal of Research in Music Education, Vol. XI, No. 1 (Spring, 1963), 75-79.
- Roby, A. Richard, "A Study in the Correlation of Music Theory Grades with The Seashore Measures of Musical Talents and the Aliferis Music Achievement Test," Journal of Research in Music Education, Vol. X., No. 2, (Fall, 1962) 137-142.
- Tarrall, Vernon V. "An Investigation of the Validity of the Musical Aptitude Profile," Journal of Research in Music Education, Vol. XIII, No. 4, (Winter, 1965), 195-206.
- Taylor, Crowin H. "Techniques for the Evaluation of Musical Status," Journal of Research in Music Education, Vol. XI, No. 1 (Spring, 1963), 55-62.
- Turner, Colin. "Examinations Should Test Course Objectives," The P.T.A. Magazine, 45, No. 5 (December, 1964), 19-24.
- White, Adolph. "The Aliferis Music Achievement Test as a Predictor of Success in Music Theory at St. Olaf College," Journal of Educational Research, 54, No. 8, (April, 1961), 315-317.

Books

- A Glossary of Measurement Terms, California Test Bureau, Del Monte, Research Park, Monterey, Calif.

- Adams, Georgia, Sachs; Torgerson, Theodore L; and Wood, Ernest R., Measurement and Evaluation for the Secondary-School Teacher, New York: The Dryden Press, 1957.
- Anastasi, Anne. Psychological Testing, New York: Macmillan Co., 1954.
- Buros, Oscar Krisen (ed.) Test in Print, Highland Park, New Jersey: The Gryphon Press, 1961.
- _____. The 1940 Mental Measurements Yearbook, Highland Park, New Jersey: Mental Measurements Yearbook.
- _____. The Sixth Mental Measurements Yearbook, Highland Park, New Jersey: The Gryphon Press, 1965.
- Chauncey, Henry and Dobbin, John E. Testing: It's Place in Education Today, New York: Harper and Row, 1963.
- Cronback, Lee J. Educational Psychology, New York: Harcourt, Brace and Co., 1954.
- Educational Testing Service. Shortcut Statistics for Teacher-Made Tests, Princeton: Educational Testing Service, 1960.
- Green, John A. Teacher-Made Tests. New York: Harper and Row, 1963.
- Greene, Harry A., Jorgenson, Albert N., and Gerberich, J. Raymond, Measurement and Evaluation in the Secondary School, New York: Longmans, Green and Co., 1959.
- Jordan, A. M. Measurement in Education, New York: McGraw-Hill, Book Co., Inc., 1953.
- Lundin, Robert W. An Objective Psychology of Music, New York: Ronald Press Co., 1953.
- Remmers, H. H., and Gage, N. L. Educational Measurement and Evaluation. New York: Harper and Bros., 1955.
- Schoen, Max. The Psychology of Music, New York: Ronald Press Co., 1940.
- Seashore, Carl E. The Psychology of Music, New York: McGraw-Hill Book Co., Inc., 1938.
- Stanton, Hazel. Measurement of Musical Talent, Studies in the Psychology of Music, Vol. II, New York: University of Iowa Press, 1935.
- Super, Donald E., and Crites, John O., Appraising Vocational Fitness. Revised Edition. New York: Harper and Brothers, 1962.
- Thomas, R. Murray, Judging Student Progress, New York: Longmans, Green and Co., 1954.

- Thomas, R. Murray, Individual Differences in the Classroom, New York: David Mckay Co., Inc., 1965.
- Townsend, Edward A. and Burke, Paul J. Statistics for the Classroom Teacher, New York: The Macmillan Co., 1964.
- Watkins, John Goodrich. Objective Measurement of Instrumental Performance, New York: Bureau of Publishers, Teachers College, Columbia University, 1942.

Unpublished Material

- Busch, Brian Richard, "An Exploratory Study of the Musical Learnings of University Freshmen Whose Sole Music Education in the Senior High School Consisted of Choral Classes," Unpublished M.M. Thesis, University of Oregon, 1963.
- Cowles, Clifton Volney, "Aesthetic Judgment of High School Music Students," Unpublished Ph.D. dissertation, University of Southern California, 1963.
- Fluke, John Holman, "The Construction, Validation, and Standardization of a Test in Music Perception for High School Performance Groups," Unpublished Ph.D. dissertation, Colorado State College, 1963.
- Hagan, Lawrence M. "A Survey of Choral Music Grading in High Schools of Five Hundred Students or More in the State of Washington," Unpublished Masters Thesis, University of Montana, 1962.
- Motycka, Arthur, "A Report on the Results of an Experimental Course of Study in General Music," Unpublished Ph.D. dissertation, University of Illinois, 1965.
- Rainbow, Edward L. "A Pilot Study to Investigate Constructs of Musical Aptitudes," Unpublished Ph.D. dissertation, State University of Iowa, 1963.
- White, Adolph Peter. "The Construction and Validation of a Group Test in Music Reading for Intermediate Grades," Unpublished Ph.D. dissertation, University of Minnesota, 1963.

Appendix I

Recommended Standardized Aptitude and Achievement Tests

Although there are a considerable number of standardized music tests available, only those which the author would recommend for high school use are listed here. Anyone wishing specific information about these or other tests should see Tests in Print or The Mental Measurements Yearbook edited by O. K. Buros. (See "Method of Procedure".)

Part A

Standardized Aptitude Tests

1. Musical Aptitude Test: Series A.
2. Seashore Measures of Musical Talents, Revised Edition.
3. Wing Standardized Tests of Musical Intelligence.
4. Musical Aptitude Profile, by Edwin Gordon (Houghton Mifflin Co., Boston)

Because this test has been developed relatively recently (1965), it had not been described, at the time of the preparation of this paper, in The Sixth Mental Measurements Yearbook (1965). Consequently, there follows a brief description of the test.

Four unique aspects of the test are: (1) It contains original music examples. (2) Performances are by professional musicians only, (3) The violin and cello are the only musical instruments used (4) It includes preference tests.¹

It measures aural perception, kinesthetic musical feeling, and musical expression. The musical expression section is divided into three

¹Harrell, Vernon V. "An Investigation of the Validity of the Musical Aptitude Profile," Journal of Research in Music Education, Vol. XIII, No. 4, (Winter, 1965), p. 195.

parts: (1) tonal imagery, which tests in the areas of melody and harmony, (2) rhythm imagery, which tests tempo and meter, and (3) musical sensitivity, which tests phrasing, balance, and style. Creativity is also appraised.

There are seven tests in all, with directions on tape. A multiple-choice answer sheet is employed. Norms are available for each grade level. The reliability coefficient is .94, while the validity coefficient is .75. ". . . the reliability coefficients of the test are about as high as those generally reported for academic aptitude and diagnostic achievement tests."²

"The Musical Aptitude Profile has one major purpose: to act as an objective aid in the evaluation of students' musical aptitude so that the teacher can better provide for individual needs and abilities."³

5. Drake Musical Aptitude Test.

Part B

Standardized Achievement Tests

1. Beach Music Test.
2. Knuth Achievement Tests in Music.
3. Kwalwasser-Ruch Test of Musical Accomplishment.
4. Watkins-Farnum Performance Scale: A Standardized Achievement Test for All Band Instruments.

Reasons for the omission of other tests from this list are: (1) The test is out of print, (2) The test is outdated, (3) The test is too advanced, (4) Either the test has unsatisfactory coefficients of validity or reliability, or else there is no available information in this regard.

²Gordon, loc.cit., p. 54.

³Ibid., p. 52.

Some observations regarding the tests which have been listed are¹¹

(1) Each test normally comes with a teacher's manual and student's answer sheets. (2) Although some can be administered from the piano, most come with records or tapes. (3) Validity coefficients range from .60 to .87, which are minimal to satisfactory. (4) Reliability coefficients range from .64 to .94, which are highly satisfactory.

Appendix II

The following chart represents an attempt to indicate graphically the tests, types of tests and techniques which are suggested for the evaluation of various educational objectives within three specific areas of high school music.

Part A

General Music

<u>Educational Objective</u>	<u>Objective Nature</u>	<u>Subjective Nature</u>
History	<u>Beach Music Test</u> , True-False (T-F) Completion (C) Multiple-Choice (MC) Matching (M)	Essay Oral Reports (Written) (Oral) Short Answer (S-A)
Theory	<u>Beach Music Test</u> <u>Kwalwasser-Ruch Test</u> <u>of Musical Accomplish-</u> <u>ment.</u> T-F, C, M-C, M.	Assignments
Appreciation (Listening)	T-F, C, M-C, M, Recognition	Essay Oral S-A
Composition		Writing Performing
Form	T-F, C, M-C, M.	Essay Oral Assignments S-A
Style	T-F, C, M-C, M.	Essay Oral Assignments Reports (Oral) (Written) S-A

Development of taste		Observation Oral Reports
Choral Music		
Sight Singing	T-F, C, M-C, M.	Performance Test - Individual - Group Observation
Performing	Attendance Extra Activities	Observation
Daily Class Work	Point System - Promptness - Care of Music - Memorization - Attendance - Section Leaders - Accompanists - Librarian - Robe Custodian - Private Lessons	Point System - Cooperation - Behavior - Student Directing - Interest
Development of Musicality		Observation Performance Test
Vocal Technique (and knowledge)	T-F, C, M-C, M.	Performance Test - Individual - Group Observation Essay S-A
Instrumental Music		
Sight Reading	<u>Watkins-Farnum</u> <u>Performance Scale</u> T-F, C, M-C, M.	Performance Test - Individual - Group Observation
Performing	Attendance Extra Activities	Observation
Daily Class Work	Point System - Promptness - Care of Music - Attendance - Section Leaders	Point System - Cooperation - Behavior - Student Directing - Interest

- Music Librarian
- Uniform Custodian
- Private Lessons
- Extra Practice
- Care of Instrument

Development of
Musicality

Observation
Performance Test

Instrumental Technique
(and knowledge)

Watkins-Farnum
Performance Scale
T-F, C, M-C, M.

Performance Test
- Individual
- Group
Observation