

Pittsburg State University

## Pittsburg State University Digital Commons

---

Electronic Thesis Collection

---

Summer 7-1937

### AN EVALUATION OF TESTS OF MUSICAL TALENT

Herbert Clark LeVan

*Kansas State Teachers College*

Follow this and additional works at: <https://digitalcommons.pittstate.edu/etd>



Part of the [Education Commons](#), and the [Music Education Commons](#)

---

#### Recommended Citation

LeVan, Herbert Clark, "AN EVALUATION OF TESTS OF MUSICAL TALENT" (1937). *Electronic Thesis Collection*. 70.

<https://digitalcommons.pittstate.edu/etd/70>

This Thesis is brought to you for free and open access by Pittsburg State University Digital Commons. It has been accepted for inclusion in Electronic Thesis Collection by an authorized administrator of Pittsburg State University Digital Commons. For more information, please contact [mmccune@pittstate.edu](mailto:mmccune@pittstate.edu), [jmauk@pittstate.edu](mailto:jmauk@pittstate.edu).

AN EVALUATION OF TESTS  
OF MUSICAL TALENT

A Thesis Submitted to the Graduate Division in  
Partial Fulfillment of the Requirements for the Degree  
of Master of Science

By  
Herbert Clark LeVan

PORTER LIBRARY

KANSAS STATE TEACHERS COLLEGE  
Pittsburg, Kansas  
July, 1937

## TABLE OF CONTENTS

CHAPTER	Page
I. INTRODUCTION.....	1
Origin of the Problem.....	1
Statement of the Problem.....	2
Importance of the Problem.....	2
Procedure.....	3
II. REVIEW OF THE LITERATURE RELATED TO THE PROBLEM	4
Importance of Finding the Best Tests of Musical Ability for Use in Elementary Schools.....	4
Conflicting Viewpoints.....	8
Relationship of Intelligence and Test Results.....	11
Improvement by Practice and Training.....	15
Development with Age.....	19
Methods of Measurement.....	20
Validity and Reliability of the Seashore Measure of Musical Talent.....	24
The Best Tests for Prognosis.....	29
Reliability of the Kwalwasser-Dykema Music Tests.....	33
Criteria for Selecting Tests.....	36
III. EXPERIMENTAL INVESTIGATION OF FOUR SELECTED BATTERIES OF TESTS.....	42
Description of the Tests.....	43
Administration of the Tests.....	46
Organization of the Results.....	47
Interpretation of the Results.....	48
IV. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	64
Summary.....	64
Conclusions.....	78
Recommendations.....	80
BIBLIOGRAPHY.....	89

## ABSTRACT

The problem consists of an evaluation of tests of musical ability to discover which tests are most reliable and useful for elementary school purposes.

In a review of the literature related to the problem an attempt was made to present different viewpoints fairly, to summarize evidence of outstanding tests, to analyze contradictory evidences for possible solution of the questions involved, and to set up criteria for the selection of the best tests.

For further evidence, four selected batteries of comparatively new tests were given to over one thousand children in grades two to six in the schools of Parsons, Kansas. Only a brief summary was given of the results from three of the batteries, but more space was given to summarizing some significant results from the statistical study of the other battery, namely the Kwalwasser-Dykema Music Tests.

Three new approaches were used in the attempt to discover which are the most reliable tests in the Kwalwasser-Dykema battery. The first approach was an analysis of all available reliability coefficients reported by different investigators for each of the ten tests in the battery. The second approach involved assembling fourteen sets of means for different ages, grades, nationalities, and races. These averages were analyzed and compared with results from fifteen different groups of Parsons' schools. The third



approach was a study of the accuracy with which the raw scores made on each test of the battery and other tests or parts of tests were predictive of the total scores made on the whole battery of the Kwalwasser-Dykema Music Tests.

In the third approach, the purpose was not only to discover evidences as to the reliability of the tests but also to find out which brief tests actually gave the most information about the musical ability of elementary school children. The criterion of musical ability used in this was the total scores which were assumed to be the best available index of the musical ability of an individual, or at least the most practical method of ascertaining the quarter in which an individual should be classified as to musical ability.

The tests finally selected are ranked according to their probable value for use in elementary schools as follows:

1. Drake Test of Musical Memory, Forms A and B.
2. Kwalwasser-Dykema Tonal Movement.
3. Kwalwasser-Dykema Tonal Memory.
4. Seashore Pitch.
5. Seashore Tonal Memory.
6. Kwalwasser-Dykema Rhythm Imagery.
7. Kwalwasser-Dykema Pitch Imagery.

## LIST OF TABLES

TABLE	Page
I. RANK ORDER OF THE AVERAGES OF ALL AVAILABLE RELIABILITY COEFFICIENTS REPORTED BY DIFFERENT INVESTIGATORS FOR THE KWALWASSER-DYKEMA MUSIC TESTS.....	34
II. RANK ORDER OF THE MEANS OF THE TOTAL SCORES AND OF THE MEANS FOR EACH TEST OF THE KWALWASSER-DYKEMA MUSIC TESTS MADE BY THE DIFFERENT GRADES IN WHITE SCHOOLS AND IN A NEGRO SCHOOL AT PARSONS, KANSAS.....	51
III. RANK ORDER OF THE TESTS IN THE KWALWASSER-DYKEMA BATTERY ACCORDING TO THE ACCURACY WITH WHICH THEY PREDICTED THE MEANS OF THE TOTAL SCORES MADE BY EACH GRADE OR GROUP IN PARSONS SCHOOLS.....	52
IV. RANK ORDER OF AVERAGES OF THE MEANS MADE ON THE KWALWASSER-DYKEMA MUSIC TESTS BY PARSONS SCHOOL CHILDREN AND THOSE REPORTED BY OTHER INVESTIGATORS.....	56
V. RANK ORDER OF HIGHEST CORRELATIONS FOUND BETWEEN TOTAL SCORES ON THE KWALWASSER-DYKEMA MUSIC TESTS AND VARIOUS OTHER TESTS OR PARTS OF TESTS.....	60

## CHAPTER I

### INTRODUCTION

#### Origin of the Problem

Proposed testing in Parsons, Kansas. In planning improvements to be made in the course of study in music for the elementary schools of Parsons, Kansas, the question arose regarding the advisability of giving tests of musical aptitude to all children in grades two to six as an aid to a better guidance program in music.

Reports from Rochester. For seven years the elementary schools of Rochester, New York have used tests of musical ability as a regular part of a musical guidance program. Reports of benefits derived suggested the value of following their example. (24. pp. 223-236)

Testing may be expensive. Extensive batteries of available tests require much time, expensive equipment, good conditions, skill and care in giving and scoring tests, and competent judgment in interpreting the results. (19, 36)

Danger of futile efforts. Warning has been published that there is danger of much futile work, loss of time and money in the extensive use of standard tests of musical talent by unskilled investigators. (12. p. 196)

Possibility of adequate brief tests. These considerations suggested the idea that there might be some adequate, inexpensive, brief tests, easy to give, with quite useful results which would be simple to interpret.

#### Statement of the Problem

The problem consists of an evaluation of tests of musical talent to discover which brief tests are the most reliable and useful for elementary school purposes. The problem requires consideration of the advisability of using tests of musical talent in the elementary schools; a study of the needs, benefits, conflicting viewpoints, and issues involved; examination of the methods of measurement, analysis of statistical reports, and setting up of criteria for the evaluation of tests; and experimental trial of selected tests, interpretation of results, and evaluation of tests according to the criteria.

#### Importance of the Problem

The need of convenient tests. Many children in the elementary schools are being neglected in musical training or are not getting the training best suited to their needs. This is largely because we have no convenient and reliable way of measuring their potential musical capacities as a basis for better musical guidance.

Great differences in ability should be better known, in order that proper provision may be made for musical activities best suited to individual needs. Therefore, the musical aptitude of elementary school children should be tested earlier, more accurately, quickly, and impartially.

The problem is vitally important to all elementary school pupils and their parents, and to all the teachers, music supervisors, administrators and others who are or should be interested in the welfare of children, and active in providing enrichment and maximum, well-balanced development for them in the most desirable phases of living.

#### Procedure

Library, statistical, and experimental techniques were used in studying the problem. The literature related to the problem was reviewed, results of statistical studies reported by different investigators were carefully analyzed, and findings from this analysis were compared with results from an experimental trial of four selected batteries of tests which were given to over one thousand children in the elementary schools of Parsons, Kansas. Tests were evaluated according to criteria set up for the selection of the best tests of musical talent for use in elementary schools.

## CHAPTER II

### REVIEW OF THE LITERATURE RELATED TO THE PROBLEM

#### Importance of Finding the Best Tests of Musical Ability for Use in Elementary Schools

Importance to pupils who may specialize in the study of music. Musical talent needs to be found earlier in life than aptitude for other professions. Instrumental technique needs to be acquired early while the muscles are growing and pliable. Very long training and great expense are usually necessary to perfect technique and to become familiar with even a small portion of the vast store of fine musical literature. Talented youth, if undiscovered and untrained, may fail to produce many beautiful compositions that might have been recorded for the benefit of mankind. Some of the very best music in the world has been written by very young composers.

Serious deficiency in one or more of the most essential musical capacities may handicap an individual in attaining musical success. Disappointment, failure, and great loss of time and money may be the fate of those who have the desire but not the talent to succeed in music. (40)

Importance to pupils who may not specialize in the study of music. The majority will find their life work in other activities rather than in music, but no doubt

their lives need to be enriched by suitable musical training, perhaps as good consumers rather than as extensive producers of music. A better understanding of the musical ability of this great majority, revealed through good tests, would facilitate provisions for appropriate musical activities better suited to individual differences.

The gifted few, who specialize in music, probably should not be favored unduly over the majority with respect to the opportunity for realizing happiness through proper musical training.

Importance to teachers, music supervisors, administrators and parents. A favorable solution of this problem of testing for musical talent promises light on a question that often puzzles them, namely, how to better judge the musical ability of children:

Through the use of these tests the powers and weaknesses of the student in the essential matters upon which musicianship and musical achievement are based are clearly, accurately, quickly, and impartially disclosed. (17)

Thus far, relatively little use has been made in elementary schools of tests of musical talent:

However, at present we find relatively little use made of these tests by music teachers for the selection of talent. This probably due to a lack of an adequate appreciation of them. (26. p. 324)

Yet the elementary school is probably the place where tests of musical talent could be of greatest service and value:

A considerable number of studies at the college level have been presented. While the results of the aptitude tests at this level have considerable value in guidance work as a directive influence, it seems that as a selective agency the use of tests at this point cannot compare in value with their use before special instruction begins in the public schools. (26. pp. 318-319)



Tests as an aid to customary ability grouping for singing. It is a common custom for teachers to seat pupils in the music period according to their opinion of the pupils singing ability. The best singers are seated in the back seats. The average singers are seated in the middle seats in each row. The poorest singers are seated in the front seats. Placement of individuals may often be largely arbitrary guess-work, without adequate justification except perhaps that of expediency. Limited time and skill may prevent better appraisal of the talent of many individuals.

Tests of musical talent should aid teachers to supplement personal judgment with more objective measurements, to classify individuals more accurately, and to justify their decisions with more tangible evidence than mere opinion.

Tests as an aid to better music teaching. Proper testing should aid in the evaluation of activities, materials, methods and procedures in teaching. Proper testing if done systematically and recorded for whole groups should aid teachers and music supervisors in checking the efficiency of their teaching. Groups of relatively equal ability should show relatively equal gains in a given time.

The use of measures of music talent and achievement should bring a closer understanding of the student, which thereby should induce a more efficient teaching technique. In applying test data it is important that a positive attitude be maintained. Tests and measurements should not be used as a means of keeping children away from music; rather, the aim should be to encourage the child to develop such capacities that he inherently may have. But since there are decided individual differences in the range of musical talent, it will not suffice to give the same kind and degree of instruction to all students. Rather, the teacher must be prepared to appreciate and use tests and



measurements as an indispensable aid in giving the proper kind and amount of instruction to the various types of talent. This, I believe, will bring the most efficient and successful teaching of music.  
(26. p. 324)

Tests as an aid to wiser musical guidance. What information will tests of talent be likely to disclose which will be most helpful in guiding a pupil into suitable musical activities? The information that an individual ranks very high in his total score may indicate that he should be examined more thoroughly to justify encouragement towards specialization in music, with special opportunities and advantages.

Pupils doing poor work in regular music classes may be studied objectively to find out if there are natural deficiencies that cause failure. A pupil ranking very low could be tested further to discover if possible the causes of his apparent deficiencies, to remedy them so far as practicable or to direct him into musical activities suited to his capacity and interest.

A knowledge of the fineness of pitch and rhythm discrimination may be useful in advising students about what instrument to begin to study. Fine pitch discrimination is necessary for high achievement in playing a string instrument. Anyone with a poor sense of pitch should not be encouraged to play the violin, or cello, but may be advised to play the piano or other similar instrument that does not depend on the player for accurate pitch.

The same opinion may be justified in perhaps a less degree regarding the playing of wind instruments such as the flute, clarinet, saxophone, cornet and trombone, which may be blown out of tune by players with a poor sense of pitch. Pupils who are poor in sense of pitch, but good in sense of rhythm, may be advised to study percussion instruments or dancing.

We need to know more nearly the exact degree of the musical aptitude of children in the elementary schools in order to determine as nearly as possible before training begins the chances pupils have of attaining a certain level of achievement. Proper advice and guidance may then be given as to the kind and amount of music study they need so that adequate provision may be made for activities best suited to their individual differences and ample opportunities and advantages offered to the extent their ability justifies. The aim is to help the teacher to accomplish in the relatively limited time available the most that is possible by the wise exercise of whatever talents the pupils have.

### Conflicting Viewpoints

Heredity and environment. Watson, chief exponent of one view, thinks musical talent is almost entirely the result of environment and proper training. (51) Seashore, chief representative of the opposite view, believes it is

a gift bestowed very unequally upon individuals. (38. p. 6)

Mursell and Glenn reviewed the literature on this subject and decided that there was not much known with certainty regarding the inheritance of musical ability. (33)

Farnsworth summed up the literature and concluded as follows:

....one is practically forced to conclude that the question of musical heredity is still a moot one. While genetics can illuminate the issue to some extent, I doubt that its contributions can ever be conclusive. With human stocks it is difficult to make very reliable Mendelian counts. And, even where musicality can be demonstrated as prevalent in certain family lines for generations, how can the factor of social inheritance be completely separated from that of biological inheritance? (7. p. 245)

Isolation of capacities. Authorities seem to agree that musical talent is a complex combination of capacities. Seashore maintains that musical talent is a hierarchy of at least twenty-five inherited capacities, which may exist in various proportions in an individual. There are also very great differences between individuals. One child may be two hundred times as keen in pitch discrimination as another equally bright child. Similar but not quite so large differences are found for tonal memory and other capacities. Seashore explains that these differences are covered up in ordinary musical performances. He contends that each factor must be isolated for accurate measurement. (40. pp. 449-452)

Other authorities believe that the most essential abilities in music are so complex in actual use that they

cannot be separated into their different elements for measurement without destroying their musical value and any measurements of these separate elements are not valid measures of actual musical ability. (29)

Fixity of capacities. Seashore believed that his phonograph tests measured elemental capacities that have relative fixity. He states that a test accurately measures these fixed capacities to the extent that the scores made on the test do not vary with intelligence, nor improve with practice and training, nor show a development of the capacity with age. (38. p. 65) These points are still questionable.

Basis of musical talent. Is musical talent a special capacity or merely a specialization of general ability? Probably in the past almost everyone has regarded musical talent as a special inherited capacity. But those who think that musical talent is largely due to the influence of environment and training on general ability have a different view. Mursell contends that general intelligence and ability when wholeheartedly applied to musical effects results in what we call musical talent. The following is a sample of his conclusions:

Musical power is associated with general traits of mind, cultural interest, personality, and physique. The typical musical personality certainly should not be the narrow specialist....musical ability is not a special, inherited talent but a specialization of general ability....Musical ability is simply the ability to deal with structure embodied in tone. In this respect it is precisely analogous to every other kind of mental ability....All mental ability is

capacity to deal with structural elements and to give an organized response. Musical talent is simply this general capacity running in a special channel and working through a special medium. What, then determines whether a man shall be a musician or not? I am not inclined to deny that some hereditary factors may affect the situation, but, in general, the significant forces would seem to be the direction of interest and will. To become a musician means the concentration of general ability in a special medium. (32. pp. 369-370)

Mursell and Glenn discussing Thorndike's (47) theory of correlation, according to which high ability in one field is a probable index of high ability elsewhere, make this comment:

It now seems likely that every individual has one or two fields where his ability is highest, and that he shades off from these by slow gradations to fields where he has but little talent. (33. p. 19)

They reviewed several early studies made in Europe which show that musicality and high intelligence go together, and decided that the American studies use too narrow a base, since dependence is entirely upon results from the Seashore tests, while the European findings are more true to fact because the investigators used a broader, more functional base as their criterion of musicality.

#### Relationship of Intelligence and Test Results

The following quotation from Mursell and Glenn is a very significant statement regarding the relationship of intelligence and test results:

There may be no relationship between Seashore test performance and intelligence, and still be a very close one between functional musical ability and intelligence. As a matter of fact, this is precisely our own opinion. We feel that there is ample justification for stating that musicality goes with high intelligence. (33. pp. 19-20)

Research reports of positive correlations of intelligence with musical talent. Stanton reported that in the ten year program of testing for talent at the Eastman Conservatory of Music, involving over one thousand students, there was a positive correlation between scores on the Iowa Comprehension Test and the scores on the Seashore tests. (43) This is interesting because it contradicts former findings of other investigators in other cities. That there is probably some real basis of relationship between intelligence and musical talent has since been verified and measured by Tilson, who reports a study made of the records from 1924 to 1935 of college freshmen at the Indiana State Teachers College, Terre Haute, Indiana. Scores made on the Seashore tests by the freshmen were compared with their scores on intelligence tests:

The scholarship index in ear training and sight singing rises sharply as the talent scores and psychological ratings rise, and within a group having the same talent the scholarship index varies with the psychological ratings.

It can be predicted that the average scholarship index in ear training and sight singing of the students with both music talent and psychological rating in the highest quarter will be 91.

It can be predicted that one hundred per cent of the students with music talent and psychological rating in the lowest quarter will make grades in ear training



and sight singing below the median. It can be predicted that the average scholarship index of these students will be 43. Since this is seven points lower than that required for permission to take practice teaching in the training school it seems clear that these students should be asked to withdraw from the music curriculum.

It can be predicted that almost nine out of ten students whose music talent is in the lowest quarter will make grades below the median regardless of their psychological rating.

Students with talent scores and psychological ratings below the first quartile should be strongly advised to withdraw from the college music curriculum. (49)

Mosher at San Jose State College, California, reports in 1955 that he tried to determine what entering students were the best risks for teacher training in the music department. Such factors as intelligence, sight reading, scores on certain Seashore tests, etc., were weighted for the determination of the final index, or estimate of risk. The results were highly satisfactory, because he thought they showed at the outset which students were promising. The students were notified of their chances of success early enough in their college career so that, if they wished, they might change their objective. Perhaps the music profession was saved some weak material which would not have been a credit to the profession either musically or educationally. (31)

Research reports indicate little relationship between intelligence and musical talent. Farnsworth found very little relationship for scores on the Thorndike Intelligence Test and scores on the Kwalwasser-Dykema Music Tests for

sophomore psychology students at Stanford University. In reporting this Farnsworth states that Chadwick compared the scores made on the Kwalwasser-Dykema Music Tests with scores made on the American Council Psychological Examination by sixty-seven music majors at the Colorado State Teachers College. The correlation coefficient was  $.26$ , plus and minus  $.08$ , which is slightly higher than Farnsworth's findings.

Farnsworth in the same report also states that Test compared the scores made on the Thurstone Intelligence Test and the total scores on the Kwalwasser-Dykema Music Tests by one hundred seventy-five music students at Syracuse University. The correlation coefficient was  $.03$ , indicating no relation whatever. (8. pp. 62-63) In a later study Farnsworth found that intelligence tests give a somewhat better prediction of the several types of music grades than did tests of musical talent. (6)

Higsmith found the Terman Intelligence Test correlated  $.423$  with marks on courses in musical performance, while the Seashore tests, taken singly or as a whole, correlated  $.312$ . (10)

Kwalwasser states that the highest correlation coefficient he has ever obtained for intelligence and music tests were  $.34$ , made on the Otis Intelligence Test and the Kwalwasser-Dykema Music Tests by seven hundred junior high school students. In this same article Kwalwasser gives the



following correlation coefficients for the Army Alpha Tests and the Seashore Tests: Pitch .35; Intensity .24; Time .12; Consonance .06; Memory .26; and with the Iowa Qualifying Examinations, Pitch .01 to .05; Intensity .02 to .11; Tonal Memory .08 to .07. (23. p. 42)

Hollingworth found no significant relation between high intelligence above 135 I. Q. and the Seashore Tests. (11) According to Kwalwasser, Seashore himself found very little relationship between intelligence and his tests. He states that Seashore believed that:

....above the level of intelligence required to understand and execute directions of the tests (mental age of about ten years) performance in pitch discrimination, perception of intensity, perception of consonance, and tonal memory is not symptomatic of intellectual endowment. (23. p. 39)

Kwalwasser concluded:

On the whole we must conclude that general intelligence is a poor indication of musicianship and that musicianship cannot predict intelligence. (23. p. 40)

#### Improvement by Practice and Training

Relative fixity in elemental capacities. It is difficult to tell whether musical talent is capable of much increase in an individual through correct exercise, or whether it is a relatively fixed capacity that one merely learns to use more skillfully by proper practice. Seashore's teaching is that innate capacities do not increase much, if any, with training. With a little exercise they

function as well as they probably ever will. Training does not significantly increase the amount but may improve the use of these capacities. They may be compared to eyesight which may be tested for normal vision when a child is quite young. (38)

There is a relative fixity in the elements of musical ability. This fixity is relative because probably an individual rarely makes the best possible use of his full physiological capacity when the measurement is made. Physiological capacity is itself a relative term in genetic psychology. It does not justify the conclusion that if one is born deficient in some or all capacities, nothing can be done about it. Even small talents may do surprisingly well if fully and wisely exerted. Ordinary ability may become remarkable if properly trained to an extraordinary degree. At least there is always some activity best suited to individual needs. (40. pp. 450-451)

Farnsworth carefully investigated and summarized more than a dozen of the principal studies that have been made on the effect of training for both long and short periods on the different Seashore tests. No positive evidence of improvement after much training was found. (7)

Since then McCarty reported no practice effect by fifth grade children after four repetitions of the Seashore pitch and intensity tests, but marked improvement on the test of tonal memory. However, she reports that the

children had some difficulty understanding and following directions on the tonal memory test, which she thinks was too hard for them. Perhaps the improvement was due largely to better understanding of the directions and how to cope with the problems of the test. (27)

Barnard found that elementary school children with no music lessons scored lower on the Kwalwasser-Dykema Music Tests, than did those who had training. (1)

Farnsworth in another study gives a digest of his previous investigations and brings his report up to date by reporting others, but finds little conclusive evidence on the problem. He concludes thus:

In any training study the procedures should be carefully outlined. Tests are always made on trained or semi-trained subjects and so reveal abilities. The extent to which they indicate capacities is not known. Retesting and the methods so far employed in short "training" periods do not seem, in the main, to change test scores appreciably. Nothing definite is known concerning the results of long exposure to a musical environment. To show the effects of this or of formal musical training, tests must be given before and after the musical exposure. (5. pp. 373-374)

Kwalwasser states that the mean for untrained students on the Kwalwasser-Dykema Music Tests is 176.25; for students having at least ten lesson, the mean is 187.50; for the untrained fifth grade, the mean is 173.18; as compared with 177.86 for the fifth grade with at least ten lessons. For the untrained eighth grade, the mean is 176.39; as compared with 186.80 for the eighth grade with ten music lessons.

The untrained improved three points in three grades.

The trained improved nine points in the three grades. Kwalwasser concludes that training attracts chiefly the talented, but not every one trained develops musicianship. Training does not increase talent and may not increase interest. The inferior soon stop. (23)

Several of the Kwalwasser-Dykema Music Tests are achievement tests and are more susceptible to training than tests of talent. This may account for a large part of the improvement made by students with some special training.

The evidence from the literature is quite strongly in favor of Seashore's theory that fundamental elements or capacities underlying musical ability have a relative fixity and, with a little exercise do nearly as well as they probably ever will.

Probably a reasonable conclusion we may form from various opinions is that the physiological basis of musical talent is an inherited combination of capacities, which is dependent upon environment and training for its full development. It is often so responsive to training that it is hard to decide how much is inherited and how much is due to environment or training. The isolation and measurement of elemental capacities, such as discrimination of pitch, intensity, time, and rhythm, are probably not direct measures of musical talent but indirectly may add important evidences of natural aptitude, and thus may indicate the relative amount

of potential talent if other factors not measured by tests support this evidence. (9)

Development with Age

Apparent contradiction. In spite of the assertion by Seashore (38. p. 65) that there should be no significant improvement with age, he provides three distinctly different sets of norms, one for fifth grade, another quite higher for eighth grade, and another still higher for adults. This apparent contradiction he explains is not real, but is caused by physiological and cognitive limits. An individual probably never reaches his physiological limit when the measurement is made. The test may become monotonous, tiresome, or uninteresting. Then it is a test of cognitive limits. Adults are likely to approach more nearly to their physiological limits than eighth-grade students. Likewise eighth-grade students are more likely to approach their physiological limits than fifth-grade pupils. (38. pp. 51-55) This does not harmonize with the fact that the average mental age of the eighth-grade student is about the same as an adult.

Hollingworth found that forty-nine children, with an I. Q. above 135, obtained about the same mean on the Seashore tests as average children, and yet their median intellectual ability was that of an adult. (11)

DeGraff findings contradict Seashore's norms:

It is usually believed that age influences rhythmic discrimination; that the older the person is, the better his rhythmic discrimination will be, but such is not the case, as a glance at the distribution table (his own) will show the distribution curve of the three age groups, namely, fifth-grade, eighth-grade, and adult, reveals an unusual degree of similarity. Although the adult has the advantage of maturity, enabling him to comprehend and concentrate on the test, his score on the rhythm test is only slightly higher than that of the fifth and eighth grade school child. (7. p. 239)

Kwalwasser evidently believes likewise since he states regarding the norms for his two tests:

It will be observed that the norms make no allowance for differences in either age or training, which means that these factors do not make for significant differences in scores. (21)

Farnsworth in his study of the Seashore-Kwalwasser Tests concluded:

In the present writer's study the fifth-graders displayed median percentiles on the melody and harmony tests of 60, 41; the eighth-graders, 75, 61; and the adults, 68, 69. No definite conclusion can be made from these data, however, since the tests are quite unreliable, and a change of one point in the raw score alters the percentile ranking by as much as ten points.

At the present writing one can state only that, while age differences are found in the Seashore battery, the causative factors underlying these phenomena are not obvious. (5. p. 336)

#### Methods of Measurement

Three common methods. The method used in most schools is the estimate of the elementary school teacher who may not be able to judge very accurately and is further handicapped by the limited time available for music. There



is hardly time enough for a little group singing which seems to be the most suitable musical activity in the elementary school. Thorough investigation of the musical talent of individuals is prohibited by lack of time, skill, and facilities.

Another method may be the estimate of a music supervisor who goes from room to room throughout the school system to help in whatever musical activity seems to be most urgent. There are usually so many demands for the music supervisor's help that little time is left to adequately investigate the musical talent of individuals. Added to the handicap of lack of time, skill and facilities, is the added handicap of not knowing the pupils very well, nor their background, case history, and personality traits.

A third method is the estimate of a private teacher of instrumental music, who specializes in a narrow field of music and may be prepared to think only of qualifications for his own instrument, and who is not likely to risk much loss in a precarious business by advising affluent pupils not to study his instrument.

Of these three, the private music teacher has the best opportunity during many individual lessons to evaluate the artistic quality of a pupil's efforts. If the teacher has a fine artistic taste and judgment his opinion is perhaps the best index of a pupil's talent.

Before this opinion may be wisely given, there must

be much practice, many lessons, and great expense to positively determine whether the pupil has the necessary combination of qualities in ample amount and proper proportion to succeed in some special type of musical endeavor.

Even then much depends upon personality traits and probable opportunities for success. The love of music and desire to work at it make a great difference. One with ordinary talent may love music so much and work so hard that in spite of his handicaps he may succeed very well not as a great performer, but as a good teacher for beginning and less advanced pupils.

The scientific method. Seashore maintains that musical talent can be measured accurately only through a scientific approach. Rigid laboratory technique must be used. The best results can be obtained only in a laboratory with adequate equipment, expert knowledge and skill, unhampered by expense or lack of time.

....it is possible to measure a specific capacity during performance in unhampered musical mood, and in every respect in the actual musical situation. In measuring capacity in singing at the present time, the singer performs in an acoustically treated music room in which there are no instruments other than a microphone (of which he may be quite oblivious); but from that microphone there are made simultaneously phonograph records and records of three or four cameras operating simultaneously in such a way that from these records every detail of the performance as a whole or at any moment in the performance, may be reconstructed with high order of precision. There the singer performs in the musical mood and in the musical situation, but the instruments analyze and set forth the elements involved. (40. p. 452)



Scientific methods outside the laboratory. A few of the essential measurements may be made outside the laboratory, but conditions and controls must be scientific. This means that one basic capacity at a time should be isolated for measurement and all other conditions and influences, that might affect the accuracy of the measurement, should be controlled so that the results would be reliable and could be checked by further measurements.

Conclusions should be limited to what was involved in the one factor measured. Thus, a high score in pitch discrimination does not justify the conclusion that an individual has high musical talent, but only that he has a very keen sense of pitch. This may be desirable for musical success, yet he may be very deficient in other equally desirable capacities.

Since application of the conclusions should be limited to the factor or factors measured, therefore, in proper musical guidance, the results of testing should be supplemented by more extensive examination, adequate audition, case history, consideration of personality traits, and probable opportunities for achievement. (38, 40, 41)

Phonograph records of laboratory tests. In order to make some of the scientific measurements available for popular use, particularly in schools, five of the laboratory tests were recorded on double disk Columbia phonograph records. At a later date a sixth record, rhythm, was added.

Each record tests a basic capacity. Pitch, intensity, time, and rhythm are measured in terms of least perceptible differences; consonance, in ability to judge degree of consonance and dissonance, and tonal memory in terms of memory span for a sequence of unrelated tones. These tests are named "Measures of Musical Talent."

### Validity and Reliability of the Seashore Measure of Musical Talent

Farnsworth reviewed the principal investigations of the validity of the Seashore tests and concluded that the only practical way to prove the worthwhileness of a test was to establish the degree of accuracy with which it predicted success in some worthwhile musical endeavor such as graduation from a music school. (5. p. 370)

The most extensive and reliable evidence of this kind comes from the Eastman School of Music, University of Rochester, New York, where a ten-year program of testing and retesting was reported by Stanton and Koerth. During the ten-year period, over one thousand students were rated by means of the first five Seashore tests and the Iowa Comprehension Test, into five groups, as advocated by Seashore, thus:

Best	10	per	cent:	stimulate	enthusiastically
Next	20	"	"	encourage	freely
"	40	"	"	encourage	
"	20	"	"	question	
"	10	"	"	discourage	(38. pp. 67-68)

Music grades at the conservatory, musical achievement, and the per cent graduating, correlated closely with this grouping.

Of the best 10 per cent,	60 per cent graduated.
" " next 20 " "	42 " " "
" " " 40 " "	33 " " "
" " " 20 " "	23 " " "
" " " 10 " "	17 " " "

These students were retested after three years of musical training. Only slight gains resulted from the retest. (44)

One of the most important findings is that the coefficient of contingency between the first and second administrations of the five tests combined was .61 (maximum .89) for grades 7-9, and .77 (maximum .91) for grades 10-12. The experimenters' general conclusion from their evidence is that the Seashore tests have high value as measures of musical talent. (14)

Testing in the grade schools of Rochester, New York.

In the spring of 1934, Ruth C. Larson (25) reported obvious benefits from five years use of the Seashore tests of talent for ability grouping and musical guidance in the elementary schools of Rochester, New York. The turnover of instruments, was less because of the more careful placement of instruments. The level of performance of instrumental organizations was raised considerably. The most talented students were especially encouraged and provided with special opportunities. Teachers studied their students more closely. This proved worthwhile and beneficial to both teachers and pupils. Parents relied more than formerly on prediction and guidance and requests for service required scheduling of this work weeks in advance.

From these measures and other available data adequate information may be acquired, which will enable the music psychologist to classify the prospective music student, and then prescribe the nature of training best suited to him.

We have found that with our present test material this is successful in groups as low as the late fourth A grade in school, and from that point on up to and including adults.

As the results of the talent tests show such decided individual differences, we must recognize that all children are not born equal in this respect, and we must attempt to provide suitable instruction for these various levels of talent. (25)

Criticisms of the tests. Severe criticisms against this apparently scientific approach have been made by Moos, (29) Farnsworth, (5) Mursell, (32. pp. 373) and others. Their criticisms ranged from unreliability, and lack of validity, to condemnation of the whole scheme as thoroughly fallacious. Gehrkens gives a good summary of the attitude of many musicians towards the whole testing movement.

No single part of a test, then, is reliable as an indication of musical talent. In fact, musicians in general have been highly skeptical concerning the entire group of tests, and have frequently taken the attitude that although the Seashore "Measures of Musical Talent" may offer some clue to an individual's inborn musical ability, yet a person might pass all the tests successfully and yet be lacking in some of the most important elements of musical talent. In other words, there is no agreement among musicians that the tests are valid for determining the complete complex referred to as musical talent. They are undoubtedly valid for certain phases of such talent, but they do not test the person's ability to work in the musical medium, their failure at this point probably being due to the fact that they depend too greatly upon measuring certain items in isolation instead of observing

them in combination with others. In music, the whole is much more than merely the sum of its parts. (9. pp. 212-213)

Defense of the tests. The music teacher wants an accurate wholesale judgment. Seashore simply says that certain facts are known and may have a certain bearing on the problem, that the value of selected measures hinges on their basic character, and that the musical guide must not depend solely on the results of the tests but must supplement them with adequate audition, case history, personality traits, and probable opportunities. Other criticisms he answers thus:

....all scientific measurements must be specific in order that the conditions may be controlled, the results recorded and the experiment repeated.

This type of measurement is fully in accord with the now generally recognized principle that any given factor must ultimately be evaluated in its natural setting as a whole, so that the outcome of the functioning of capacities is ultimately evaluated in relation of the total personality to the total situation. Behaviorism and Gestalt psychology at their best thus deal with specific factors. (39)

It is true that talent is not the sum of specific talents. It is an integrated whole. But we get a truer and deeper insight into the integrated whole by employing the scientific method of fractioning: that is, by observing one aspect at a time. (40. pp. 451-452)

Conclusions. A consensus of the opinion of competent musical educators as revealed by this review of the literature is that educational or vocational advice and decisions should not be given solely from test results. Many other factors not measured by available tests must be thoroughly investigated before results of testing can be effectively

utilized in musical guidance. Even advice based on all this evidence should be made with due reservations.

Some scientific methods of testing may be suitable for the elementary school, if they are brief and convenient, and if they are as valid and reliable as other good educational tests.

In the elementary school, especially, we need tests of a more musical nature, that really measure the actual musicality of children, that are easier to administer and more reliable and useful in results.

Achievement cannot be predicted very accurately because so much depends upon circumstances, influences, and opportunities. Pupils will always be variable in interest, attention, effort, concentration, and in freedom from distracting or disturbing elements in every testing situation.

Exhaustive diagnosis of the ability of individuals is probably beyond the scope of the elementary schools. To discover and measure all the strong or weak points, and to find the causes of difficulties of individuals may require thorough examination and diagnosis by skilled specialists in a laboratory. Probably the elementary school can usually succeed only in a broad classification of pupils into the more musical, the average, and the less musical. A quick estimate and prescription is usually necessary because time is limited and there are so many pupils needing immediate help of various kinds.



As an addition to other essential information, good tests of musical talent will re-enforce and supplement the teacher's judgment by helping to decide the extent of each capacity measured, to foretell better, than without them, the probabilities of attaining a certain level of musical achievement, and to select the best and the poorest pupils in a short time. These tests may also enable one to learn more about some pupils than would be learned otherwise in months, to discover outstanding ability, to diagnose some difficulties or defects, to evaluate teaching, and to motivate learning.

#### The Best Tests for Prognosis

Review of many studies. Grace Van Dyke More reviewed nine studies of the Seashore tests and found they showed that Tonal Memory and Pitch had the highest coefficients for reliability and validity, but decided they had prognostic value only to a limited degree. From her own experiments she concluded that:

Physical, mental and emotional conditions may alter results but indications are that there is a very real relationship between results of certain types of music tests and the student future success in music study....among the highest correlations are the tests that endeavor to set problems of a general musical situation rather than to isolate and measure a single factor of musical ability. (29)

Lamp chose the Seashore Tonal Memory and Pitch tests as being best of the Seashore Tests for use in an experiment to discover whether aptitude for special instruments could

41456

be determined by these tests. The results obtained caused him to decide that they did not succeed in indicating aptitude for special instruments. (24)

Highsmith reported highest reliability, .82 for Tonal Memory, and .76 for Pitch of the Seashore Battery. (10) Drake also found these two tests most valid and reliable of the Seashore battery by obtaining the coefficients .94 and .84. (4) Salisbury and Smith, after one year of experimenting to discover the best tests to predict success in college sight singing, selected the Seashore tests of Tonal Memory and Pitch. They used only these two Seashore tests and one of their own construction, throughout the rest of a four year experiment, which indicated positive correlation between scores on these tests and success in sight singing. (42)

Taylor found the Kwalwasser-Dykema Pitch Imagery the best out of sixteen standardized tests for predicting achievement in sight singing and dictation. (44) But Tonal Movement is highest in reliability of any test in the Kwalwasser-Dykema battery, and Tonal Memory is next, according to the reliability coefficients reported by different investigators as follows: Bowman, (8. p. 80) Tonal Movement retest .83; Tonal Memory retest .72. Farnsworth, (8. pp. 57-60) Tonal Movement retest .73, odd-even, stepped up .85, .81, .78; Tonal Memory retest .55, odd-even, stepped up .68, .59, .58. Whitley, (53. p. 745) Tonal Movement odd-even,



stepped up .83; Tonal Memory odd-even, stepped up .55. Sanderson, (36) Tonal Movement retest .57, .39; Tonal Memory retest .51, .34. Drake, (4) Tonal Movement upper-lower half, stepped up .85, .73; Tonal Memory upper-lower half, stepped up .57, .55.

The highest reliability coefficient found in the literature was .93 for the Drake Test of Musical Memory. This was reported by Drake, (4) who also gives .85 as the result from another group. Correlation between Form A and Form B of the Drake Test of Music Memory for grade schools was .77; for junior high .73; for senior high .71. (3)

For both forms together it consistently produced a validity coefficient more than six times its probable error with different kinds of groups and unrelated teacher estimates. The coefficient of validity ranged from a  $r$  of .499 plus and minus .075 with age and training partialled out, to a raw validity coefficient of .671 plus and minus .054. (3)

Boswell selected the Drake Test of Musical Memory as the best test of musical aptitude to use in his investigation. (2) Sward selected as the best tests of musical aptitude for use in his investigation the Drake test of musical memory, Kwalwasser-Dykema test of tonal movement, and the Seashore tests of tonal memory, pitch and intensity. Sward reported raw self-correlations for the total Drake Musical Memory Test were .68 plus and minus .03 for Jewish boys, and .63 plus and minus .04 for non-Jewish boys. The self- $r$ 's, corrected by the Brown-Spearman formula, became .81 plus and minus .02, and .77 plus and minus .03. A

retest of 31 individuals in two sub tests after a month's interval yielded a coefficient of .72 plus and minus .06.

(45)

A study of the literature indicates that investigators select the tests that have the highest validity and reliability, when they wish to use only a few tests to obtain the best indication of musical aptitude.

Since high correlations were found for the more musical type of tests it seems reasonable to conclude that tests which present problems of a general musical situation may be better for practical purposes in the elementary school than tests that isolate and measure a single factor of musical ability.

The Drake Test of Musical Memory has the highest recommendations as this type of a test. It is not only of an interesting musical nature but its author has reported well established validity and unusually high reliability. As yet this has not been thoroughly tried out and outside the author's report only one other was found. While it is very good, it does not seem to show as high reliability as the author claimed.

The Seashore tests of Tonal Memory and of Pitch are apparently the most valid and reliable of the Seashore battery, but they are of an unmusical nature. This is a serious objection from the standpoint of progressive music education in elementary schools.

Tonal Movement and Tonal Memory have the highest reliability of the Kwalwasser-Dykema battery according to the investigations that have been made thus far.

Pitch Imagery proved the best out of sixteen tests for predicting sight singing ability, but it is undoubtedly an achievement test requiring special training to be of much value. According to reports its reliability otherwise is very low, but this may be due to the fact that very few pupils learn to recognize the notation of pitch as they should. This test may have great teaching value for this desirable skill.

There is need for further analysis to bring some order out of all the contradictory reports by different investigators as to the reliability of the different tests in the Kwalwasser-Dykema battery.

#### Reliability of the Kwalwasser-Dykema Music Tests

Table I shows the rank order of the averages of all available reliability coefficients reported by different investigators as for the Kwalwasser-Dykema Music Tests.

There are thirteen sets of reliability coefficients for the whole battery. Sets number one, three, seven, and ten are incomplete. If they had been complete the rank order of the tests might have been changed somewhat, but probably not much. These sets are ranked from highest to

TABLE I

RANK ORDER OF THE AVERAGES OF ALL AVAILABLE RELIABILITY COEFFICIENTS REPORTED BY DIFFERENT INVESTIGATORS FOR THE KWALWASSER-DYKEMA MUSIC TESTS

Method of computation*	Averages for the whole battery	reliability coefficients for each test										
		Tonal Movement	Tonal Memory	Time Discrimination	Rhythm Imagery	Quality Discrimination	Intensity Discrimination	Rhythm Discrimination	Pitch Discrimination	Pitch Imagery	Melodic Taste	
1 odd-even	.66	.85	.55			.66						.61
2 retest	.61	.83	.72	.53	.58	.65	.63	.47	.65	.70		.35
3 odd-even	.55	.73	.57			.39						.40
4 " "	.53	.88	.56	.56	.62	.56	.65	.55	.52	.02		.36
5 " "	.52	.78	.68	.68	.61	.54	.53	.68	.11	.48		.11
6 " "	.50	.85	.63	.63	.20	.56	.60	.28	.63	.33		.28
7 " "	.47	.83	.55	.48		.42	.43	.43	.39			.22
8 " "	.46	.81	.58	.60	.52	.44	.54	.51	.18	.36		.00
9 " "	.45	.61	.16	.64	.59	.63	.86	.55	.44	.91		.04
10 " "	.43	.85	.59	.64	.50	.48	.37	.31	.00	.53		
11 retest	.38	.55	.73	.42	.40	.53	.10	.21	.05	.42		.53
12 " "	.25	.57	.34	.19	.31	.20	.07	.27	.34	.14		.10
13 " "	.23	.39	.51	.11	.37	.08	.12	.04	.38	.28		.06
averages	.46	.73	.55	.49	.47	.47	.43	.37	.33	.33		.27

\* all odd-even computations were stepped up by the Spearman-Brown formula.

lowest according to the averages or means for the whole battery of tests. This column, marked M, is on the left side of the table. Each test is ranked according to the average of all the reliability coefficients reported for each test. This column of averages runs across the bottom of the table.

The method of computation for each set of coefficients is shown as derived from a retest or by the odd-even, split-half method, stepped up by the Spearman-Brown formula.

Sets number two, four, five, six, eight, nine, ten, and eleven, were reported by Farnsworth. (8. pp. 57, 59, 60, 68) Sets number one and three were reported by Drake. (4) Sets number twelve and thirteen were reported by Sander-son. (36) Set number seven was reported by Whitley. (53. p. 745)

As judged by these coefficients none of these tests are reliable enough to be used as the sole basis for the musical guidance of individuals. However they may be very useful for the purposes of the elementary school, namely to aid in quickly discovering outstanding talent or extreme inferiority and to group children for musical activities best suited to their needs. This may require only a rough estimate such as deciding in which quartile a pupil probably belongs.

According to these averages Tonal Movement is the only test with even a moderate degree of reliability for individual

prognosis. Tonal memory is far behind and not much above Time, Rhythm Imagery, and Quality, but we should not form an adverse opinion solely on this evidence. Perhaps much of the apparent imperfection is due to lack of adequate understanding and skill in using the tests and interpreting results, and there may be important teaching values involved.

#### Criteria for Selecting Tests (35, 48)

The criteria commonly agreed upon by authorities, as the best for selecting tests is as follows: validity, reliability, objectivity, administration, applicability, manual, norms and standards, and equivalent forms. Four other items probably should be added because they seem especially desirable for evaluating tests of musical ability to be used in elementary schools. These four items are: musicality, interest to children, teaching value, and brevity. Each item of the criteria will be briefly considered in the above order.

Validity. Validity is the degree to which a test measures what it is claimed to measure. Validity involves the question of the general worthwhileness of a test, and the benefits that may actually result from its use. In the final analysis the best judgment of experts determines the validity of a test. They may use various aids in forming their judgment. The only really essential statistical technique is that which refines or assists in obtaining the best consensus of the judgment of experts. If other means of validation such as teachers marks



are used, validity coefficients should be given by single grade or age ranges as well as by total ranges.

Reliability. Data on reliability should include the exact number of cases of each kind and the methods by which they were chosen. Individual grade-ranges, and single age-ranges, when available, should be presented. The method of computing the reliability coefficient must be clearly set forth, whether on alternate forms of the test, repetition of single forms, or the split-halves method corrected by the Spearman-Brown formula. The significance of the difference between the reliability of the author's test and similar tests should be computed, as well as the probable errors of estimate of the test under consideration.

The effect of obtaining reliability coefficients for a single grade has, perhaps, not been sufficiently taken into account in drawing conclusions regarding the reliability of certain tests.

....the greater the self-correlation of a test the more reliable it is. The reliability coefficient is, however, not an entirely satisfactory measure of reliability, for it is affected by the distribution, in the trait measured, of the particular group studied.

To secure a reliability coefficient of .40 from a group composed of children in a single grade is probably indicative of greater, not less reliability than to secure a reliability coefficient of .90 from a group composed of children from second to twelfth grades. The spread of talent is four times as great in eleven grades as in a single grade. The correlation in the second case would have to be .914 in order to indicate as close a relationship as that shown by a reliability coefficient of .40 in a single grade. (16)



This statement by T. L. Kelley is worth investigating. It may make the coefficient of reliability appear better for some tests. It does not agree with statements by other authorities as follows:

Per cent of predictability for specific reliabilities. (13)

Reliability of criterion	Per cent of predictability
.99-----	.90
.98-----	.86
.95-----	.78
.90-----	.68
.85-----	.61
.80-----	.55
.75-----	.50
.70-----	.45
.60-----	.37
.50-----	.29
.40-----	.23
.30-----	.16

The following quotation seems to indicate still another opinion of interpretation.

The coefficient of alienation for a correlation of .30 is .95, which means that the error of prediction is approximately 95 per cent as great as the error of pure chance. (54)

Objectivity. The best tests usually have clear cut test situations definite and adequate method of pupil responses, provision for specific time limits, and unequivocal answers.

Administration. Ease of giving, scoring, tabulating, and interpreting results are very important to consider. Some music tests require great skill and carefulness to ensure a fair degree of reliability in results. Tabulation is necessary if the fruits of comprehensive testing are to

be realized. Even where the teacher is not prepared to interpret many of the implications of the test results, she should not feel that a test is complete unless blanks for tabulation have been provided, properly filled out and filed as permanent records.

Applicability. The actual application that will be made of the results should be determined. Many times nothing is done of any use or benefit with the results after they have been obtained. Understanding should be clear as to whether the purpose is classification, diagnosis, prognosis, or all of these to some specific degree. Definite provision should be made for systematic utilization of results to obtain full benefits from them.

Manual. The manual should provide in a comprehensive manner information on the following considerations: The purposes, construction, validation, and reliability of the test; directions for giving, scoring, and tabulating test results; directions for norms and standards for interpretation of the results; and suggestions on how the results may be used for purposes of classification, diagnosis of difficulties or defects, motivation, and application of the test results in musical guidance.

Norms and standards. There should be adequacy of sampling, age, and grade norms, percentiles for age and grade series, norms for sub-sections, and norms for different groups with which the test will be used.

Equivalent forms. Equivalent forms are often desirable. If they are available it is important to consider the method of obtaining equivalence, intercorrelations, average scores, standard deviation, and range of scores by grade or age groups for the various forms.

Musicality. From evidence in the literature of the subject we have drawn the conclusion that isolation of basic capacities is not always necessary to obtain comparatively reliable and useful estimates of musical talent. Therefore it is important to select the most musical type of tests for their helpful influence in developing appreciation of good music, and to avoid unmusical tests that may become distasteful to young children.

Interest to children. Other things being equal, the best results will probably come from tests that appeal to the interests of children and that involve enjoyable musical experiences.

Teaching value. In the elementary school, time for musical activities is usually very limited. Therefore it is very desirable that tests of musical ability and musical achievement be justified by their teaching value alone, entirely aside from their testing value. Many aspects of music may be effectively brought to the attention of children through the proper type of tests, and there is no doubt about the effectiveness with which the right kind of tests can be utilized to motivate the learning of many different things.

Brevity. Tests should not take up more time than is necessary to effectively utilize their important teaching values and to ensure adequate reliable measurements. Children may soon become tired or inattentive and measurements then become less reliable. Time for testing should not unduly encroach upon other equally important activities which compete for attention in an already over-crowded schedule.

Other considerations. There may be other elements involved in selecting tests for specific situations as for example:

1. Expensive facilities may be required, such as a special testing room, a well-tuned piano, an expert pianist, a good phonograph, and many clerical assistants for scoring and recording results.

2. An expensive musical guidance program may be necessary to ensure actual application or systematic utilization of the results to obtain full benefits from them, otherwise extensive testing surveys may not be justifiable.

3. Scheduling testing services may be difficult, because other important activities will probably interfere with convenient scheduling, especially if very many tests or much time is required.

Only general criteria considered. While some special considerations should be thoroughly investigated for specific situations, our present study is limited to general criteria for selecting tests of musical ability for use in elementary schools.

## CHAPTER III

### EXPERIMENTAL INVESTIGATION

#### OF FOUR SELECTED BATTERIES OF TESTS

New tests to be investigated. The Seashore tests have been extensively investigated for many years and Pitch and Tonal Memory quite well established as the most valid and reliable of the Seashore battery. Yet, because they are unmusical, monotonous, rather long and somewhat difficult, and because time for experimental testing was limited, these tests were not given. It seemed more desirable to investigate newer, more interesting tests. Two batteries of tests were selected because they seemed to come nearest to satisfying the demands of the criteria. These tests were the Drake Test of Musical Memory and the Kwalwasser-Dykema Music Tests. The McCreery Rhythm and Pitch Test and the Kwalwasser-Rueh Test of Musical Accomplishment was given also, for reasons that will be explained later.

The Drake Test of Musical Memory seemed to be rather well standardized and its validity and reliability better established by its author than is usual for tests of musical talent. Its reliability may not be as high as claimed by its author, but it probably has good teaching value as it is quite musical.

The Kwalwasser-Dykema Music Tests have been greeted favorably as an improvement in many respects over the old

Seashore tests, or at least a step in the right direction. (28, 53) Yet the apparent unreliability reported for some parts of this battery may be a serious handicap. (8)

The McCreery Rhythm and Pitch Test, which is an unstandardized talent test by C. L. McCreery was also included, not because it came nearest to meeting the criteria, but, to see how it compared with tests whose validity and reliability was better known. A thousand blanks for this test were furnished free by a musical instrument company which uses this method of discovering prospective buyers for their instruments.

#### Description of the Tests

The McCreery Rhythm and Pitch Test has two parts. In the first part, "Rhythm Recognition", two measures, are tapped by the administrator, then silence exactly one measure, followed by tapping the same two measures exactly as at first or tapping them differently, as printed changes occur in administrator's copy which the students cannot see. Pupils write in the proper square on their test blank the letter S for same or D for different. There are ten of these answers to be made.

The second part, called "Pitch Recognition", is played on the piano by the administrator. There are five sections in this part of the test, with eight answers to be made in each section. All the exercises, except some brief melodies, are three measures long. The first and third measures may



be the same or different and are always separated by a silence of exactly one measure. This part of the test begins with two notes in a measure for the first trial, then three notes, and increases in length. In one section called "Single Notes" the listener must decide whether in the third measure the tones are higher or lower than the first measure and mark H for higher or L for lower. The same is true of another section called "Chords". There is no manual, only the administrator's printed exercises and the test blanks. The test is simple to administer but seems rather easy for a real test.

The Drake Musical Memory Test consists of twenty-four original two-measure melodies to be played on the piano by a good pianist. Each one of these melodies serves as a standard of comparison with four possible variables, change of key, change of time, change of note, or the same melody repeated. Only one kind of change is made in any one variable. The answers are written K for key, T for time, N for note, S for same. The first melodies are easy and have only two variables to be answered. Progressive difficulties are introduced until the more complicated melodies have seven variations to be answered. The manual has clear explanations and directions, age norms for both boys and girls, and graphs showing percentage of students receiving each score for all ages. Form A and Form B seem to be well matched, but Form A is at a disadvantage because pupils

may not readily grasp all the complications involved until they have had much practice.

The Kwalwasser-Dykema Music Tests consist of ten parts on five double disk Victor phonograph records. They are intended to test ten elements of musicality by a total of two hundred seventy-five samples.

Tonal Memory, Pitch Discrimination, Intensity Discrimination, Time Discrimination, and Rhythm Discrimination may be compared to the Seashore tests but are briefer. In this investigation they will be called the five talent tests.

Quality Discrimination, Tonal Movement, Melodic Taste, Pitch Imagery, and Rhythm Imagery, measure elements that are probably more influenced by training therefore in this investigation they will be called the five achievement tests.

One large double faced blank with columns of numbered squares makes it easy for young pupils to write S for same, or D for different, in the correct square. For some parts of the test different letters are used. This may cause young children to write a wrong answer when they could have written the correct answer. Some unreliability may be due to this factor. Much care must be taken to drill young children in writing plainly and consistently the letter changes that are needed. Some of the children are almost sure to write some of their letters in the wrong column. The administrator must keep checking up on all the pupils, especially at the beginning of each part of the test, to

see that the right columns and squares are being used. The record helps in this by announcing the number which is printed beside the correct square.

Four matrices are supplied to facilitate scoring. The manual contains clear explanations and directions, the notation of what is played by the record, a scoring key with correct answers and percentile norms or standards.

The Kwalwasser-Ruch Test of Musical Accomplishment is well known. It consists of ten parts that are intended to test knowledge of musical notation. Time limits of this investigation prevent detailed comparisons that might be made, and as this test does not directly enter into our problem further description will be omitted. (18)

#### Administration of the Tests

The Kwalwasser-Ruch Test of Musical Accomplishment was given to two hundred forty-nine children in the fifth and sixth grades as a survey of their knowledge of musical notation.

All the other tests were given to over one thousand children in grades two to six, to twenty-six eighth and ninth grade students in a drum corps, and to one hundred forty Negro students in grades three to eight in a Negro school.

All testing was done in the regular class room of each grade. Rooms were well lighted and ventilated and quiet as possible. It was explained to the children that a series

of tests would be given and that the scores they made would not be used for grades, but only to help the investigator find out which was the best test or game. Their full cooperation was solicited in this project. The attitude of the investigator was that of playing a game to see who could be most prompt and accurate in writing the correct answers. The children seemed to maintain this attitude quite well.

Each part of a test was explained in simple language. Demonstrations were given on the blackboard, and pupils were asked to call aloud in unison the judgments for the first few trials. Some drills were given in locating proper columns and squares for writing the answers. Opportunity was given for asking questions and some of the children were questioned to make sure that all fully understood what was to be done. Before starting the actual test the directions were stated as given in the manual for the test.

Only one test battery was given at a time. In some cases it was necessary to give parts of a battery at different times, rather than all at the same sitting. There was usually an interval of a week between different tests.

#### Organization of the Results

On the front page of the Kwalwasser-Dykema Music Tests blanks was copied for each pupil all the scores he made on all the other tests. This was for convenient reference and comparison. Percentile graphs on the front page of these

blanks were made out for each pupil to simplify comparisons. In the space above the graphs, teachers marks in music and the other subjects, for each year, were copied and also the results of three standardized general achievement tests.

### Interpretation of Results

Only significant results from the statistical study of the four batteries of tests will be given much consideration in this thesis. Since no significant relationship was found between the Kwalwasser-Ruch Test of Musical Accomplishment and the other tests, it will not be discussed. Attention will be largely directed to the Kwalwasser-Dykema Music Tests, but an interpretation of some other results will be summarized first.

1. No very significant relationship was found between the McCreery Rhythm and Pitch Test and other tests, except that a correlation coefficient of .52 was obtained between the McCreery total raw scores compared with the total raw scores on the Kwalwasser-Dykema battery. This indicates that both batteries tend to classify pupils in the same rank order. The McCreery Rhythm and Pitch Test is probably too easy and each part is probably too short. It was the first test given, but a large majority obtained nearly perfect scores.

2. Raw scores on the Drake Test of Musical Memory, Form A, compared with total raw score on the Kwalwasser-Dykema battery, gave a correlation coefficient of .41, but

for Form B the result was .59, probably because the second and third grade were not included. An operetta prevented giving them Form B. For grades four to six the result was .58 for Forms A and B added together as one test and compared with total scores on the Kwalwasser-Dykema battery. Pupils in grades four to six did slightly better on Form B than on Form A, probably because of better understanding and practice effect, or Form B may be easier than Form A.

3. Three new approaches were used in the study of the Kwalwasser-Dykema battery. Results from the first approach have been already summarized in Table I, page 34. Before discussing the second approach, the third will be briefly described as follows: The raw scores made by the pupils on each test in the battery were compared to three different total scores; first, to their total scores made on the whole battery of ten tests; second, to their total scores made on five so-called talent tests in the battery; and third, to their total scores made on the remaining five tests which are called achievement tests. These comparisons will be considered later on pages fifty-eight to sixty.

The second approach which we will now consider consists of an analysis of the means or averages for the battery and for each test in the battery made by fifteen different groups of children in the schools of Parsons, and a comparison of these results with a similar analysis of all



available means reported by different investigators. This involved assembling from various sources, fourteen sets of means made on the Kwalwasser-Dykema battery by groups that were very different from each other, and which represented a widerange of ages, grades, nationalities and races.

In Table II, the fifteen groups in Parsons are ranked from highest to lowest according to their averages or means of the total scores made on the whole battery listed in a column marked M near the left side of the table. The ten tests in the battery are listed in rank order from left to right according to their value or accuracy for predicting the averages made on the whole battery. The rank order of these tests was determined by computing the correlation coefficient between each column of the ten tests compared with the column marked M containing the averages of the total scores. In this computation the averages were carried out to two decimal places, which for lack of space are not shown in Table II, which does not show the relationship quite so well as if the decimals were shown also, since the range is small. The column for Rhythm Imagery compared with the column marked M gave a correlation coefficient of .98. Table III shows the results which determined the rank order of the tests in the battery.

TABLE II

RANK ORDER OF THE MEANS OF THE TOTAL SCORES AND OF THE MEANS FOR EACH TEST OF THE KWALWASSER-DYKEMA MUSIC TESTS MADE BY THE DIFFERENT GRADES IN WHITE SCHOOLS AND IN A NEGRO SCHOOL AT PARSONS, KANSAS

Race	Grade	Number of pupils N	Means of the total scores M	Means on each of the tests									
				Rhythm Imagery	Tonal Movement	Pitch Discrim.	Tonal Memory	Pitch Imagery	Intensity Discrim.	Time Discrim.	Rhythm Discrim.	Quality Discrim.	Melodic Taste
*W.	8-9	26	199	20	22	27	17	16	23	20	18	23	14
"	2-6	60	197	19	20	28	18	16	22	19	18	25	14
"	6	193	178	17	16	25	17	14	22	18	17	20	13
"	5	222	177	16	17	24	15	14	20	18	17	20	14
#N.	8	22	175	16	15	22	14	14	22	20	18	21	14
"	7	20	172	15	15	22	15	14	23	18	17	21	14
"	6	25	171	15	15	23	15	14	21	17	16	20	14
W.	2-6	911	170	15	16	23	15	14	21	17	16	20	13
N.	3-8	140	168	15	15	21	15	14	21	17	16	19	13
W.	4	187	167	14	14	23	15	14	21	16	16	21	12
"	3	170	164	14	16	22	15	14	21	16	16	20	11
N.	5	27	163	13	15	21	14	13	21	17	17	20	12
W.	2	139	158	13	16	21	14	12	20	17	15	20	12
N.	4	33	157	13	15	20	14	13	20	17	16	18	13
"	3	13	150	12	16	19	13	13	20	16	14	16	11
Averages			171	15	16	23	15	14	21	18	16	20	13

\* white schools  
# Negro school

Table III shows the correlation coefficients between the means of the total scores and the means of each test in the Kwalwasser-Dykema battery as derived from the raw scores made by 1077 children in the grade schools of Parsons, Kansas.

TABLE III

RANK ORDER OF THE TESTS IN THE KWALWASSER-DYKEMA BATTERY ACCORDING TO THE ACCURACY WITH WHICH THEY PREDICTED THE MEANS OF THE TOTAL SCORES MADE BY EACH GRADE OR GROUP IN PARSONS SCHOOLS

Means of each test	Means of total scores	Number of groups or grades	Total number of pupils	r
Rhythm Imagery	"	15	1077	.98
Tonal Movement	"	"	"	.95
Pitch Discrimination	"	"	"	.95
Tonal Memory	"	"	"	.90
Pitch Imagery	"	"	"	.89
Intensity Discrimination	"	"	"	.88
Time Discrimination	"	"	"	.86
Rhythm Discrimination	"	"	"	.84
Quality Discrimination	"	"	"	.80
Melodic Taste	"	"	"	.66

That Rhythm Imagery shows high predictive value is probably not a mere coincidence, because similar results have been obtained in other places besides Parsons, although evidently this has not been noticed before. Comparison of fourteen sets of means from various cities gives a correlation of .95 for the means of Rhythm Imagery and the means of the total scores. These results were obtained from a total of over a thousand students, age ten to adults. The means were given in the same reports as the reliability coefficients in Table I.

Another significant bit of evidence is that Pitch Imagery is close to Tonal Memory in predictive value according to these figures relating to central tendencies. As has already been stated on page 30, Taylor found Pitch Imagery the best out of sixteen tests to predict success in sight singing. Perhaps here is another test that needs to be investigated further.

The means of these tests correlate highly with the means of the total scores made by each grade or group, because the range is greatest and the different means correspond more closely with the different points on the wide range of the means of the total scores. This wide range may be due to more accurate measurement of talent, or may be due to measurement of achievement. The simplest explanation is that these tests are measuring achievement somewhat. Yet Pitch Discrimination certainly is not an

achievement test. Quality Discrimination had a range as wide as Pitch Discrimination but the means were not distributed like the means of the total scores and show a lower correlation. Consequently range alone does not explain these results.

The evidence here presented seems to indicate greater reliability for Pitch Discrimination than has previously been assumed from reports of very low reliability coefficients. In spite of some apparent defects it seems to be very good for classifying children into a few groups according to ability.

A group of twenty-six members of a drum corps in grades eight and nine were given the tests to discover if possible whether tests of musical talent would aid in the selection of applicants for membership in the drum corps. Age and grade, as well as ability give this group highest rank. The present investigation indicates that these tests will probably give good help in selecting the best and eliminating the most inferior applicants.

A group of sixty children in grades two to six was selected by music teachers as the most gifted of their pupils. Since the suggestion has been made that this is one of the most reliable methods of determining the musical aptitude of children, it is very significant to find them next in rank to the eighth and ninth grades.

The other grades and groups are ranked about as might

have been predicted from knowing them well for a long time. However, it was most surprising to find how well the white second grade did in comparison with the other grades. Many children in the second grade were quite handicapped by being very poor writers. Yet they seemed to know the correct responses about as well as any grade.

In order to obtain more certain evidence of the reliability of the findings thus far, it seemed desirable to compare Parsons results with results from other cities and also with results from college students and adults. Analysis similar to that shown in Tables II and III was made of the fourteen sets of means referred to on page 50. By adding the means together we obtain approximately the mean of the total score for the ten tests, or if we divide this by ten, by placing a decimal point, we obtain the average of the means of the battery.

Table IV shows the rank order of the means of the total scores made on the Kwalwasser-Dykema Music Test by Parsons school children as compared to those reported by the investigators listed on page thirty-five in connection with Table I. All of these different results fit together in remarkable order into a nicely graduated scale a progressive increase from the lowest grades to the highest with adults in the first rank. The Negroes are slightly below the means for the white children. This throws them out of order with the rest; but even they, among themselves,



TABLE IV

RANK ORDER OF AVERAGES OF THE MEANS MADE ON THE  
 KWALWASSER-DYKEMA MUSIC TESTS BY PARSONS SCHOOL  
 CHILDREN AND THOSE REPORTED BY OTHER INVESTIGATORS

Investigator	group	grade	N	M-10	M-5T	M-5A	R.I.
Farnsworth	adults		78-104	21.5	22.6	20.4	19
"	college musicians		50-77	21.3	21.6	19.0	20
"	" boys		108-228	21.3	22.6	20.0	20
"	Panatrops data			20.7	21.8	19.6	19
"	college girls		100-121	20.6	22.0	19.2	19
LeVan	drum corps	8-9	26	19.9	21.0	18.8	20
"	gifted children	2-6	60	19.9	20.9	18.4	18
Sanderson	Jewish	7-10	45	18.8	20.6	17.0	18
"	German	8	45	18.4	20.4	16.4	17
Farnsworth	girls	8	109	18.2	19.6	16.8	18
"	boys	8	100	17.9	19.6	16.2	17
Sanderson	Italian	8	45	17.9	20.0	16.0	16
LeVan	white	6	193	17.8	19.6	15.9	16
"	"	5	222	17.7	18.9	16.3	16
Sanderson	Negro	8	45	17.5	18.8	15.8	16
LeVan	"	8	22	17.5	19.1	16.2	16
"	"	7	20	17.2	19.4	15.6	14
Sanderson	Polish	8	45	17.2	19.0	15.6	15
LeVan	Negro	6	25	17.1	19.0	15.2	15
"	white	2-6	911	17.0	18.3	15.5	15
"	Negro	3-8	140	16.8	18.4	15.1	15
"	white	4	187	16.7	18.3	15.0	14
"	"	3	170	16.4	17.8	15.0	14
Farnsworth	boys and girls	5	188-189	16.4	17.6	15.2	14
LeVan	Negro	5	27	16.3	18.0	14.5	13
"	white	2	139	15.8	17.2	14.4	13
"	Negro	4	33	15.7	17.3	14.4	13
"	"	3	13	15.0	16.2	13.7	12

are in exactly the same progressive order from the third grade to the eighth. The apparent inferiority of the Negro and Polish groups may be due to various handicaps and not to inferior ability.

In Table IV the first column of averages is marked M-10, which indicates that this column contains the averages of the means for the whole battery of ten tests. The next column is marked M-5T, which indicates that it contains the averages of the means for the five tests of the battery, designated as five talent tests. The following column is marked M-5A, indicating that it contains the averages for the five tests of the battery, designated as five achievement tests. The last column is marked R. I. showing that this records the means for Rhythm Imagery.

Table IV shows that either the five talent tests or the five achievement tests will rank groups of all ages, nationalities, and races very nearly the same as the whole battery. Moreover, Rhythm Imagery will do the same work of classification with equal reliability.

The reliability of some of the Kwalwasser-Dykema Music Tests has been rather well indicated by a number of investigations, through retesting and self correlation methods. Our study of the means and comparisons of the averages of the means, which shows remarkable agreement in results from all investigations, gives further evidence of the reliability of the tests taken as a whole.

Therefore it seems reasonable to assume that the total score made on the Kwalwasser-Dykema Music Tests is a very concise and useful index of the musical ability of an individual. If this is true then the tests that show the most significant correlation with the total score may be the best brief tests to indicate musical aptitude, because they must be measuring nearly the same things or something closely related or they would not show a significant relationship. Kwalwasser verified this conclusion in a personal letter to the writer as follows:

If a single test is to be given with the hope that it will predict the scores on the total test, the one with the highest correlation is obviously the right one to give. But I cannot hazard a guess as to the one it would be. (22)

Since there seems to be some uncertainty as to the validity and reliability of some parts of the Kwalwasser-Dykema Music Tests, perhaps one or two valid and reliable parts might give as good or better indication of the talent of the majority of individuals as a more complete but less reliable examination. In school administration the most common use of aptitude tests and intelligence tests is for classification or ability grouping of pupils. Some brief parts of the Kwalwasser-Dykema battery will probably classify groups or grades in practically the same rank order as the total scores of all the tests together.

Over one hundred correlation coefficients were computed as to the relationship between the total scores on the Kwalwasser-Dykema Music Tests and the McCreery Rhythm and

Pitch Test, The Drake Test of Musical Memory, and the different parts of the Kwalwasser-Dykema Music Tests. Most of the correlation coefficients were quite low. However, there may be some significance in some of the higher coefficients.

Table V shows the rank order of the highest correlations found between total scores on the Kwalwasser-Dykema Music Tests and the raw scores on various other tests or parts of tests. The Kwalwasser-Dykema battery is made up of two types of tests designated as five talent tests and five achievement tests. Similarity to some of the Seashore "Measures of Musical Talent" suggested the name "five talent tests" as used in Table V. They are as follows: tonal memory, intensity discrimination, time discrimination, rhythm discrimination, and pitch discrimination. The other tests are called "five achievement tests" to distinguish them and because perhaps most of them are more influenced by training than the five so called talent tests.

Besides the Kwalwasser-Dykema tests, the Drake Test of Musical Memory, the McCreery Rhythm and Pitch Test, and the Kwalwasser-Ruch Test of Musical Accomplishment are included in Table V.

Quality Discrimination is noticeable for its absence. It is the only part of the Kwalwasser-Dykema battery to get too low a correlation to be included. No significant relationship was found between scholarship and total scores made on the Kwalwasser-Dykema Music Tests. The correlation coefficient was .39. Scholarship was represented by a

TABLE V

RANK ORDER OF HIGHEST CORRELATIONS FOUND BETWEEN TOTAL SCORES ON THE KWALWASSER-DYKEMA MUSIC TESTS AND VARIOUS OTHER TESTS OR PARTS OF TESTS

Names of tests or parts of tests correlated		grade N	r
K-D Rhythm	K-D 5 Achievement Tests	4-6 557	.79
" Pitch Discrimination	" Total Score	6 192	.69
" " "	" " "	4 142	.63
" " "	" " "	5 223	.62
" " "	" 5 Talent Tests	4-6 569	.62
" Rhythm Imagery	" Total Score	3-6 723	.62
" Pitch Discrimination	" " "	3-6 638	.60
Drake Musical Memory B	" " "	4-6 408	.59
" " " A & B	" " "	4-6 554	.58
K-D Rhythm Discrimination	" 5 Talent Tests	3-6 600	.58
" Tonal Movement	" 5 Achievement Tests	4-6 555	.58
" Time Discrimination	" 5 Talent Tests	3-6 605	.57
" Tonal Memory	" Total Score	2-6 825	.54
McCreery Rhythm and Pitch	" " "	3-6 641	.52
K-D Tonal Memory	" 5 Talent Tests	4-6 559	.51
" Pitch Imagery	" 5 Achievement Tests	4-6 555	.50
K-R Musical Accomplishment	" Total Score	5-6 249	.47
K-D Rhythm Discrimination	" " "	3-6 668	.46
" Melodic Taste	" 5 Achievement Tests	4-6 554	.45
" Time Discrimination	" Total Score	2-6 847	.44
" Intensity	" 5 Talent Tests	4-6 559	.42
Drake Musical Memory A	" Total Score	2-6 561	.41
K-D Melodic Taste	" " "	3-6 670	.40
" Tonal Movement	" " "	3-6 636	.40
Scholarship	" " "	4-6 352	.39



scholarship rating given each pupil after a careful investigation of his whole school record including teachers marks and grade equivalents from three standard achievement tests given at the end of three successive years.

No significant relationship was found between elementary school teacher's estimates of the musical talent of their pupils and the total scores on the Kwalwasser-Dykema Music Tests. Elementary school teachers were requested to make a list of their pupils arranged in the rank order of their estimate of the pupils musical talent and also to classify them into three or more groups according to their opinion of the pupil's talent.

However a very significant relationship was found between the music teacher's estimates of the musical talent of their pupils and the total scores on the Kwalwasser-Dykema Music Tests as shown by the consistently high record of the group of sixty gifted children in grades two to six.

This investigation gives us greater confidence in the reliability of these tests. From a comparison of the means and the averages of the means we conclude that whatever these tests measure, they do it quite consistently, as far as the means are concerned. The means are very nearly alike for each comparable grade or group. If this line of reasoning is correct then all of these tests have a good deal of reliability in classifying pupils into three groups of best, average, and poorest, because the means are so much alike.



-52-

This may be good evidence of the reliability of these tests, if the more gifted children consistently make higher means and conversely, the less gifted nearly always make lower means, as seems to be true in the present study. The substantial agreement of the results of the tests with what one should logically expect as a result is good evidence of a significant degree of reliability in the tests taken as a whole.

Racial and national differences. Differences in means made on the Kwalwasser-Dykema Music Tests by Negro and white children of Parsons and means reported by Sanderson for eighth grade Negro, Polish, Italian, and German children of Chicago Public Schools, and Jewish children in grades seven to ten in an orphanage may be due to age, grade, economic, social, educational, and other factors rather than race.

It is really quite surprising that there should be so little difference between grades, not considering probable handicaps of the Negroes. Table IV shows that averages are very nearly the same for Parsons' Negro grades six (17.1), seven (17.2), eight (17.5), Sanderson's Polish grade eight (17.2), his Negro grade eight (17.8), and his Italian grade eight (17.9). Parsons' white grades five (17.7) and six (17.8), slightly excel Parsons' negro grades six, seven, and eight, but this is probably due to the white children's better advantages since some of the tests may be much influenced by musical training. The apparent slight inferiority

of the Negro is probably not real. Adequate investigation would be necessary to determine the truth.

Sanderson gives a brief digest of six previous studies which compare negroes and whites. The best one by C. B. Johnson included 3300 fifth, eight grade, and college students who were given the Seashore tests. Comparing results with Seashore's norms, Johnson found no significant differences between negro and white scores. (36)

Farnsworth also gives a digest of these studies and adds several others of different nationalities, including Hawaiian, Chinese, Japanese, Korean, American Indian-White, Chinese and Japanese-White. He comes to this conclusion:

The writer would hazard this conclusion from his work on oriental students. There is nothing to indicate that oriental blood per se would predispose one to lower scores on the tests in question. It is conceivable, however, that lack of acquaintance with occidental music may cause a lowering of scores.

The data on the other racial groups are obviously too confusing to form bases for generalizations. The factors of test unreliability and unequal motivation are probably responsible for much of the lack of agreement. (5. p. 363)

## CHAPTER IV

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

#### Summary

Chapter I. The problem consists of an evaluation of tests of musical talent to discover which brief tests are most reliable and useful for elementary school purposes.

The procedures used in the solution of the problem involved the use of the library, statistical, and experimental techniques. The literature related to the problem was reviewed; significant statistical reports by various investigators were carefully analyzed; and results from different investigations were compared with results from an experimental trial of four selected batteries of tests which were given to over one thousand children in the elementary schools of Parsons, Kansas. Tests were evaluated according to criteria set up for the selection of the tests most suitable for use in elementary schools.

Chapter II. Review of the Literature Related to the Problem. Some of the findings from the literature and conclusions based thereon are summarized as follows:

1. All children need musical activities of some kind well adapted to their widely varying abilities and interests.

Convenient and reliable tests are needed to classify individuals and groups according to musical talent and

achievement in order to facilitate the provision for suitable musical activities and adequate musical guidance.

The purpose of tests and musical guidance is to help the teachers to accomplish in the relatively limited time available the most that is possible by the wise exercise of whatever talents the pupils have.

2. Differences in viewpoints may result in quite different decisions as to which are the best tests to use for specific purposes. Therefore it is necessary sometimes to decide which viewpoint is best. For example, it is difficult to tell how much of achievement is due to heredity and how much to environment. It may not be necessary to know exactly, but the question must be decided whether or not it is best, as maintained by Seashore (38), to isolate inherited or basic capacities and accurately measure them as the best practical basis for estimating and classifying the musical talent of children. In this investigation, the viewpoint is taken that isolation is not necessary, because many experimental studies, including the present one, seem to show that tests consisting of general musical problems measure talent equally well and are more interesting and useful especially for children.

3. Seashore and many others present much proof that there is no significant relationship between intelligence and musical talent. Mursell and Glenn suggest that, "There may be no relationship between intelligence and the

Seashore tests, and still be a very close one between functional musical ability and intelligence." (33. pp. 19-20) However Stanton, (43) Tilson, (49) and Mosher (31) proved quite conclusively that there actually is a significant relationship between intelligence and the Seashore tests.

4. Evidence from Seashore (38, 39, 40, 41) McGarty (27) and others indicates that basic capacities have relative fixity. Evidence from Bernard (1) and Kwalwasser, (23) contradicts this somewhat. Farnsworth reviewed experimental studies and concluded that previous studies have not been conducted so as to give conclusive evidence.

5. Progressive improvement of scores with age is found in all tests of talent as shown in Tables II and IV and by most reports of the Seashore tests. This apparent contradiction of the theory that there is no development of basic capacities by age, is explained by Seashore by the fact that an individual probably does not reach his physiological limit when the measurement is made, especially if the test becomes tiresome or uninteresting, then it is a test of cognitive limits. Therefore Seashore provides different norms for grades five and eight and for adults. Some studies report no increase with age as Hollingworth (11) and DeGraff (7. p. 259).

6. The best common estimates of musical talent are probably those of private instrumental teachers regarding their pupils, but even these should be supplemented by

scientific methods, the best of which can only be obtained in the most up-to-date laboratory. (40. p. 452) Even the best tests should not be used as the sole basis of judgment, but should be supplemented by more extensive examination, adequate audition, case history, consideration of personality traits, and probable opportunities for achievement. (41)

7. The validity and reliability of the Seashore tests have been intensively studied for many years. The most extensive favorable evidence is reported by Stanton, (43) Stanton and Koerth, (44) and Larson. (25) They agree that the tests have high value for predicting success in music study, if used in connection with a good intelligence test to help make the estimate.

Severe criticisms have been made against the Seashore tests by Moos, (29) Farnsworth, (5) Mursell, (32. pp. 373) and others. These criticisms have been answered quite well by Seashore, (4) except in one or two points, namely, that the phonograph tests are monotonous and uninteresting to children, and that these tests do not measure the total personality in the total musical situation. The latter may be done in a properly equipped laboratory, but it cannot be done in the school room with the Seashore phonograph tests.

All authorities agree that musical guidance should not be given solely from test results. The isolation and



measurement of elemental capacities such as discrimination of pitch, intensity, time, and rhythm are probably not direct measures of musical talent but indirectly may add important evidences of natural aptitude, and thus may indicate the relative amount of potential musical talent if other factors not measured by tests support this evidence. (9)

Achievement cannot be predicted very accurately, because it depends so much upon circumstances, influences, and opportunities.

Exhaustive diagnosis of individual talent or individual defects is beyond the scope of the elementary school, which can best classify pupils into broad groups of the more musical, the average, and the less musical. A quick estimate and prescription is usually necessary. Good tests will do this work quickly and will supplement the teacher's judgment, aid in better understanding of pupils, better diagnosis, prognosis, and remedial work, and better classification, evaluation of teaching, and motivation of learning.

8. The thirteen available studies of the Seashore tests agree that Pitch was most valid and reliable and that Tonal Memory was second best of the Seashore battery.

Only five studies were available of the comparatively new Kwalwasser-Dykema battery, but all five agreed that Tonal Movement was by far the most reliable and Tonal Memory the second best of the Kwalwasser-Dykema battery.

89

Only three reports were found regarding the Drake Test of Musical Memory. The author reports the highest validity and reliability that was found for any test in the literature. (3, 4) The only other statistical study of this new test that could be found was by Sward, who reported somewhat lower reliability coefficients. (45) Boswell selected this test as the best to use in his investigation but no reliability coefficients were reported. (2)

Taylor found that the Kwalwasser-Dykema test of Pitch Imagery was the best out of sixteen standardized tests for predicting achievement in sight singing and dictation, (44) but it is an achievement test requiring a knowledge of notation to be of much value. According to other reports its reliability is quite low. This may be because the pupils were guessing largely. It is well known that few pupils really have skill in the recognition of the notation of pitch. This test probably has high value for teaching this very desirable skill.

9. Conflicting reports regarding the reliability of the comparatively new Kwalwasser-Dykema Music Tests suggested evaluating their reliability from some different standpoints, rather than use the usual methods of retest, or split-half, odd-even correlation which have shown such contradictory results. Three approaches were used. First, an analysis of all available reliability coefficients reported by various investigators; second, an analysis of all available means

from these different investigators, and comparison with results from over a thousand children in the elementary schools of Parsons, Kansas; third, a study of the accuracy with which the raw scores made on each test in the battery and other tests or parts of tests predicted the total scores that were made on the whole battery. Since raw scores were not available from other investigations, the third approach was necessarily confined to the raw scores made by the children in Parsons.

Thirteen sets of reliability coefficients were assembled from various sources and averages for each test and each set were computed. Tonal Movement ranks first with an average of .74. This is nineteen points higher than Tonal Memory which ranks second with an average of .55. Other results may be studied by referring to Table I, page 34.

Results from the second and third approaches are considered in the summary of Chapter III, under number 3. They seem to indicate a higher degree of reliability than is shown by averages of available reliability coefficients.

10. The criteria for selecting tests is as follows: Validity, reliability, objectivity, administration, applicability, manual, norms and standards, equivalent forms, musicality, interest to children, teaching value and brevity. With the exception of the last four, these are the criteria commonly agreed upon by authorities as the best for selecting standardized tests. The last four items are added because

they seem especially desirable for evaluating tests of musical ability to be used in elementary schools.

Chapter III. Experimental Investigation of Four Selected Batteries of Tests. Only significant results from the statistical study of the four batteries of tests are given in this thesis. Attention is directed largely to a consideration of uncertainties concerning the Kwalwasser-Dykema Music Tests, while other results are only briefly summarized.

Although the Seashore tests have been studied for many years and Pitch and Tonal Memory well established as the most valid and reliable of the Seashore battery, yet, because they are unmusical, monotonous, and rather difficult, they were not given, as time for testing was limited. It seemed more desirable to study newer, more interesting tests, namely, The Drake Test of Musical Memory, Forms A and B; the Kwalwasser-Dykema Music Tests; and the McCreery Rhythm and Pitch Test. To survey knowledge of notation, the Kwalwasser-Ruch Test of Musical Accomplishment was given to two hundred and forty-nine children in the fifth and sixth grades. The other tests were given to over one thousand children in grades two to six in the schools of Parsons, Kansas.

1. The unstandardized McCreery Rhythm and Pitch Test was extensively studied by correlating raw scores made on

each part with the corresponding part of the Kwalwasser-Dykema Music Tests. Combination of parts were likewise compared. No significant relationship was found. However the McCreery total scores compared with the Kwalwasser-Dykema Total Scores gave a correlation coefficient of .52, which means that there was some tendency for both batteries to classify pupils in the same rank order.

The McCreery Rhythm and Pitch Test is probably too easy. It was the first test given, but a large majority obtained perfect or nearly perfect scores. Each part of the test is probably too short also.

2. Total scores made on the Drake Test of Musical Memory, Forms A and B compared with total scores on the Kwalwasser-Dykema Music Tests gave a correlation coefficient of .58; for Form B .59; for Form A .41.

The lower correlation of Form A may be explained by the fact that second and third grade children were included who probably did not understand directions or perhaps the test is too advanced for them. An operetta prevented giving them Form B. Grades four to six did slightly better on Form B than on Form A perhaps due to better understanding and practice effect, or Form B may be slightly easier than Form A.

3. Three new approaches were used in the attempt to discover which are the most reliable of the Kwalwasser-Dykema Music Tests. The first approach which has already been summarized consists of an analysis of all available

reliability coefficients, results of which are shown in Table I, page 34.

A second approach consists of an analysis of all available means reported by different investigators and comparisons of them with results from one thousand seventy-seven children in the public schools of Parsons, Kansas. Results are shown in Table II, page 51, Table II, page 52, and Table IV, page 56.

This analysis and comparison of averages shows a remarkable agreement in results from all investigations. All these different results from a wide range of ages, grades, nationalities, and races, fit together in surprising order into a finely graduated scale of progressive increase from the lowest grades to the highest with adults in the highest rank.

Although the Negroes do not seem to fit well into this scale because their scores are slightly below the scores of the white children, which is probably due to social, economic, and educational handicaps, yet the Negro grades are in consecutive order among themselves, as are also the white children when taken alone.

The conclusion seems reasonable that whatever these tests measure, they do it quite consistently, as far as averages indicate consistency and reliability.

The substantial agreement of results of the tests with what one should logically expect as a result, is good



evidence of a significant degree of reliability in the tests taken as a whole.

If the more gifted children consistently make higher means and conversely the less gifted nearly always make lower means, as seems to be true, this is further evidence of the reliability of the Kwalwasser-Dykema battery.

In the third approach to the question of the comparative reliability of the tests in the Kwalwasser-Dykema battery, three comparisons were made. First, the raw scores made by each child on each test were compared with his total score for the whole battery; second, each score was compared with his total score on five of the tests which are called talent tests on account of their similarity to the Seashore tests; third, each score was compared with the total score on the remaining five tests which are called achievement tests to distinguish them and to indicate that perhaps some or all of them are more influenced by training than the five so-called talent tests.

Two purposes were kept in mind, first to discover evidences as to the reliability of the tests, second to find out which tests actually gave the most information about the musical ability of elementary school children. The criterion of musical ability used in this was the total scores which were assumed to be the best available index of the musical ability of an individual, or at least the most practical method of ascertaining the quarter in which an individual

belongs. If this reasoning is correct the best test is the one which most nearly predicts the total scores, and classifies or ranks the individuals and groups most nearly like the total scores ranks them. This conclusion was verified by one of the authors of the tests, Jacob Kwalwasser, in a personal letter to the writer, quoted on page 58, in which the statement is also made that he (Kwalwasser) cannot hazard a guess as to which test this would be.

Over one hundred correlation coefficients were computed. Results from large numbers of children were included in these computations. Many of the correlation coefficients obtained were quite low. However, there may be some significance in some of the higher coefficients. Table V shows the rank order of the highest correlation coefficients obtained by comparing these three type of total scores made on the Kwalwasser-Dykema battery, with the raw scores made on each test in the battery, and with various other tests or parts of tests.

In Table V the results from Rhythm Imagery are of outstanding interest. The correlation coefficients are as follows: .79 for raw scores made on Rhythm Imagery compared with total raw scores made on the five achievement tests in the battery by five hundred fifty-seven children in grades four to six; and .62 for raw scores made on Rhythm Imagery compared with total raw scores made on the whole battery of ten tests by seven hundred twenty-three children in grades

three to six. These are among the highest correlations obtained, but the significant point is that Rhythm Imagery may have value for classifying pupils according to ability besides its value as a test of reading rhythm notation and its value for teaching the recognition of rhythm notation. These three outstanding values indicate that Rhythm Imagery should be studied further to see if it should be added to the list of the best tests for use in elementary schools.

An interesting comparison in Table II that gives surprisingly high correlation coefficients which may be questioned unless understood, is shown in Table III. For example in Table II the column of fifteen means for Rhythm Imagery was correlated with the column of fifteen means of the total scores. In this computation these means were carried out to two decimal places which for lack of space are not given in Table II. The result was a correlation coefficient of .98. Without decimals the result was .95 for doing the same thing in all the available results reported by various investigators which included fourteen complete sets of means made by a wide range of ages, nationalities and races.

The explanation for these high results is that in these two columns corresponding values are located on corresponding points of the scale in both ranges.

An examination of the evidence from all available sources shows that there is substantial agreement between

the findings of previous investigators and the results of these three approaches, as to which of the Kwalwasser-Dykema tests are most reliable. Tonal Movement and Tonal Memory probably should be given highest rank, but the present study of the statistics reported by these investigators also reveals Rhythm Imagery as another test that may have significant value. Rhythm Imagery ranks first in Tables II, III, and V, but the importance and extent of the three values of Rhythm Imagery should be determined by further investigation and checked by different methods, before we assume too much from this evidence of superiority.

The criticism may be made that Rhythm Imagery is not a test of talent but is a test of achievement exclusively, since it tests the recognition of the notation of rhythm, which depends upon knowledge of note values acquired only by training. But for practical purposes it is difficult to tell how much musical achievement is due to training and how much is due to innate capacity. According to Kelley, achievement and intelligence are ninety per cent the same thing. (15) This is probably true also of talent and achievement. Therefore it seems reasonable to use the test of Rhythm Imagery to supplement tests that, perhaps more exclusively, measure innate capacities.

## Conclusions

A review of the literature related to the problem revealed six outstanding tests and the present investigation revealed one more namely Rhythm Imagery that probably should be added to this list. These seven tests may be ranked in three groups in the order of their probable value for use in the elementary schools. Within each group they are ranked in the order of highest reliability. If only one test is to be given the first would be the best and others may be added as desired. The first group is as follows:

1. Drake Test of Musical Memory, Forms A and B.
2. Kwalwasser-Dykema Tonal Movement.
3. Kwalwasser-Dykema Tonal Memory.

This group of three tests is given first place because: The use of these tests is justified by their teaching value alone, aside from their testing value. They consist of a few brief musical problems suitable for teaching various aspects of music. In the elementary schools especially these tests with musical value are more desirable than unmusical tests, even though the unmusical tests may have slightly higher reliability in measuring basic capacities. Basic capacities may not be as important in music as general abilities. (32. pp. 369-370) However these tests in group one rank among the highest for reliability according to the large majority of reports by different investigators. Several investigators recently selected these tests as the best tests of musical

aptitude to use when they desired to give only a few brief tests to investigate musical ability. This seems to be convincing evidence that they are the best available tests. Second choice, or the next best group of tests of musical aptitude for use in the elementary schools is as follows:

4. Seashore Pitch Discrimination.
5. Seashore Tonal Memory.

These two tests are given second place because they rank with the highest in reliability, but they do not present problems of an interesting musical nature. They consist of many monotonous repetitions of unmusical sounds, which are entirely unrelated to anything a child associates with enjoyable music. Yet where the most accurate measurement of pitch is necessary the Seashore Pitch Discrimination test is undoubtedly far superior to any other pitch test available. The Seashore Tonal Memory test may also be most accurate in measuring the basic capacity or capacities involved although this is not quite so certain. Third choice is as follows:

6. Kwalwasser-Dykema Rhythm Imagery.
7. Kwalwasser-Dykema Pitch Imagery.

These tests are given third place because they are probably the best objective tests of the recognition of the notation of rhythm and pitch for pupils who have some knowledge of notation, and they probably have excellent teaching value.

Pitch Imagery according to Taylor (46) was the best out of sixteen tests for predicting success in sight singing and ear training.



Of the Kwalwasser-Dykema tests Rhythm Imagery is probably the best single test to use for classifying pupils according to ability, because, out of fourteen tests used in the present investigation and fourteen sets of results reported by different investigators, Rhythm Imagery was the best for predicting the total scores and averages that were made on the whole battery of ten tests in the Kwalwasser-Dykema series.

### Recommendations

Further investigations should be made of these tests to verify or change the conclusions of the present study. Probably other methods of evaluation should be used also. The tests designated in these conclusions should be given at the beginning and at the end of each school year, in connection with an adequate musical guidance program which may be outlined as follows: (40)

A musical guidance program. A musical guidance program should be organized to give cooperative service throughout the entire school system. The plan should be devised so as to better adjust musical activities to fit individual needs and differences. The purpose of the plan should be to keep each pupil at his natural level of successful achievement.

(40. p. 449)

The accomplishment of this aim will be greatly aided by the proper combination of five important factors which will

be briefly considered separately. These factors are: surveys of general intelligence, musical talent, and musical achievement; effective follow-up services and classification according to musical and general ability; homogeneous grouping for special, intensive musical training; adequate cumulative records, conveniently available throughout the school system; and wise musical guidance in suitable musical activities under inspirational leadership.

1. Surveys of talent and achievement should begin as early as possible in the grades, probably, not later than the second year and should be carried on throughout the entire system as far as possible for musical guidance to be practicable.

Tests of talent and achievement should be given at the beginning and the end of each year for several reasons. Talent tests need to be checked for reliability by a retest at the end of the year. This retest may also be of service in enrolling pupils for special summer music study. Achievement tests need to be repeated to check progress during the year to see if it is proportionate to talent. The benefits derived will amply repay for the small amount of extra time, work and expense involved.

A suitable intelligence test should also be given, because this information is necessary in making out a musical talent index as an aid in musical guidance. This index of musical talent may be modeled after the plans reported by

Tilson (49) described on page twelve, Mosher (31) described on page thirteen, or the plan described on pages twenty-four and twenty-five as used in Rochester. (25, 26, 43, 44)

Proper facilities are necessary for best results. A quiet testing room, a good phonograph, and a well-tuned piano are indispensable. Administrating, scoring, tabulating and interpreting results should be in charge of someone experienced in giving standardized educational tests.

The tests designated in the conclusions of this investigation are of such a character that their use in these surveys will be fully justified by their teaching value alone in making pupils more conscious of the existence and significance of specific aspects of music. They can be applied before expensive training begins and can serve as an aid and guide in the choice of the most suitable musical activity.

A survey does not guarantee that a pupil who rates high is musical, but it furnishes a series of significant facts that will aid the musical guide, who must be a scientist in analysis of musical capacities and have a keen insight into the nature and laws of the development of the musical mind. Probably the safest method for the musical guide to use is to encourage the pupil to reveal and perhaps follow his natural urge, and then analyze the situation in the light of the success of his efforts compared to his record on the tests and determine whether the pupil's choice is wise and in accordance with his natural talent.

All junior and senior high school and college musical organizations should make these tests a part of entrance requirements. The best groups should refuse entrance to those who rank in the lowest quarter in scholarship, intelligence, and talent. But provision should be made for suitable musical activities in which the least talented can participate profitably if they desire to really work in music.

2. Effective follow-up services are indispensable. Surveys should not be attempted unless there is adequate provision for effective follow-up services for both positive and negative results. Positive results will be shown by the classification of pupils according to talent and achievement and the provision for musical activities suited to their individual needs and differences. Negative results will be shown by economical and protective services. Special individual music lessons are expensive in school or outside of school. Taxpayers should be assured that this expense is justified by adequate results in proportion to expense. Musical organizations are expensive. Economic resources should be conserved and the best interests of the school, the music director, and the music teacher should be protected by refusing to recommend the lowest talent and intelligence for expensive training that will probably show little beneficial results. To reduce the budget and get the most for the money expended, use early discovery of

talent, systematic record of achievement, and assurance that merit will be rewarded. This will be a stimulus that will bring out the best in the musical life of the school.

3. Homogeneous grouping is often desirable for some musical activities. It may not be necessary for musical activities of a general appreciational nature. But for special intensive music study pupils should be grouped according to ability if it is practicable to do so. Under keen competition, talented pupils have more incentive and do better. When grouped with the less talented they soon find out that they can do better than the average without much study, effort, or practice and they are often inclined to loaf. This may not always be true of some musical groups or organizations where a few gifted leaders can help the rest along, for there the gifted players get most of the practice and instruction and the others, frequently, are not given much attention unless they, by accident, get too much in the way, in which case they get "bawled out." The less talented soon realize, very correctly, that they will never be able to learn so easily and quickly as the gifted. Some become unnecessarily discouraged and quit altogether, while others become a nuisance by discovering that they can get reflected glory by crowding in where they are incapable of doing their part properly.

4. Adequate cumulative records are, probably, the best aids for musical guidance. Perhaps the most valid and

truthful basis for predicting achievement in the junior high school is a combination of factors from the elementary school record, including age, grade-progress, attendance, teachers marks in music, scores, dates, and names of intelligence tests, musical talent tests, musical achievement tests, and standardized tests of general achievement and scholarship, interpreted in terms of percentile rank, grade-placement scores, and amusical talent index.

Comprehensive tabulation blanks should be provided and carefully filled out as completely as possible with at least the minimum essentials required for proper musical guidance.

For trained pupils, perhaps, it would be helpful to supplement these records by a summary of significant evidences of talent, achievement, rank in contests, and degree of excellence in public performances or auditions as estimated by competent judges. If practicable, no doubt, informal cumulative records would be helpful for ready reference in teaching as well as guidance, if they contained specific evidences of the degree of talent exhibited from time to time in daily work.

5. Musical activities under inspirational leadership and suited to individual needs and differences should be provided. The most talented pupils should be given special opportunities and advantages. If they cannot afford private lessons or an instrument, they should have first chance for school-owned instruments and free lessons, providing they



maintain progress proportionate to their greater talent. If progress does not equal talent, find the cause and inform the pupil that to get free advantages he must progress according to his ability.

The large majority of average individuals who do not plan to specialize in music should be trained, perhaps, more as good consumers rather than as extensive producers of music. Suitable activities may include a reasonable amount of singing, playing, rhythmic interpretation, creative endeavor, and listening lessons. Provision should be made for ample facilities, motivation, and participation to make all children feel that music is an enjoyable form of expression through which an important part of their personality may develop even though they do not desire to make it more than an avocation. There should be a social situation in which the pupils are encouraged to live in an atmosphere of self-expression and service in music.

Diagnostic and remedial work should be undertaken as far as practicable for all who may profit by it. Those who sing or play out of tune, time, or rhythm, or have faulty tone quality may be improved by competent training if they have adequate basic capacities, which can be determined largely by fairly reliable standardized tests of basic musical capacities.

Scientific methods will act as a very strong stimulus in remedial work. The Tonoscope invented by Carl E. Seashore

and sold by C. H. Stoelting Company, Chicago , has been found to remedy faults of pitch and quality, because the pupil can see the extent of his success or failure in trying to reach the standard. Further experiments should be made with it to determine the advisability of its use and the extent of its serviceability.

Individuals and groups should have opportunities for auditions, similar to radio try-outs, in which systematic observations are made by competent judges. This gives good motivation and also an opportunity for making informal cumulative records of specific evidences of the quarter in which an individual should be classified according to talent and achievement. These cumulative records will greatly facilitate musical guidance.

The classroom teacher can have an important part in this in many ways, especially by furnishing leads or clues for the skilled auditor to unobtrusively observe and record his judgment. If properly done this early recording of specific abilities exhibited in performance may be particularly valuable in the primary grades as a basis for future musical guidance.

Parents should consult a musical guidance expert and consider with him the results of tests of talent to supplement advice of music teachers before purchasing an expensive instrument. In the light of all available information regarding the child, suggestions may be made as to the kind of instrument for which he is best equipped.

The most common test of musical talent is achievement. Children can be given much musical training and watched for achievement. This is no doubt a good procedure, but may be often wasteful of time and money.

One of the best tests of achievement is the competitive contest, where each student or group competes in his own classification, and is ranked according to excellence by competent judges. However there is some difference of opinion among authorities as to the ultimate benefits to the mass of young children from a contest in which the winner takes all and the loser who may be nearly as good suffers ignominious defeat. Serious failure is not considered a good thing for young children by many authorities. An important responsibility of the educator is to create situations in which the mass of children can celebrate well-earned successes, that will encourage them to keep trying to improve.

The musical guidance thus far seriously considered is largely educational and avocational. If an extensive and ambitious career is contemplated, a thorough analysis should be made if possible in a laboratory to evaluate the probabilities of success in the chosen course of development and especially to discover and evaluate any serious shortcomings that may exist.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

1. Barnard, B., "A Study of Seashore's Measures of Musical Talent." Music Supervisors Journal, XVII, (May, 1933), 54-56.
2. Boswell, reported by S. F. Bienstock in "Report of National Survey of Experimental Projects in Music Education." Music Educators National Conference, Yearbook, (1936), p. 278.
3. Drake, R. M., Manual of Directions, for Musical Memory Test, Bloomington: Public School Publishing Co., 1934, p. 5.
4. Drake, R. M., "The Validity and Reliability of Tests of Musical Talent." Journal of Applied Psychology, XVII, No. 4, (August, 1933) 447-458.
5. Farnsworth, P. R., "An Historical, Critical, and Experimental Study of the Seashore-Kwalwasser Tests." Genetic Psychology Monographs, IX, No. 5 (May, 1931) p. 336-374.
6. Farnsworth, P. R., "Are Music Capacity Tests More Important Than Intelligence in the Prediction of the Several Types of Music Grades?" Journal of Applied Psychology, (June, 1935), p. 350.
7. Farnsworth, P. R., "The Effects of Nature and Nurture on Musicality." National Society for the Study of Education, 27th Yearbook, 1928, part 2, pp. 239-245.
8. Farnsworth, P. R., "The Kwalwasser-Dykema Music Tests as Psychological Tools." Genetic Psychology Monographs, XV, No. 1 (January, 1931), 1-94.
9. Gehrkens, K. W., Music in the Grade Schools, Chicago: Silver, Burdett & Co., 1934, pp. 215-220.
10. Highsmith, J. A., "Selecting Musical Talent." Journal of Applied Psychology, XVIII, (1929), 486-493.
11. Hollingworth, L. S., "Musical Sensitivity of Children Who Test Above 135 I. Q." Journal of Educational Psychology, XVII, (February, 1926), 95-109.
12. Hubbard, George E., Music Teaching in the Elementary Grades, Chicago: American Book Co., 1934, p. 196.

13. Huffaker, C. L., "Predictive Significance of the Correlation Coefficient." Journal of Educational Research, XXI, (January, 1930), 47.
14. Jones, Vernon, and R. H. Brown, "Educational Tests." Psychological Bulletin, XXXI, (1934), 487.
15. Kelley, T. L., Interpretation of Educational Measurements, Chicago: World Book Co., 1927, pp. 193-213.
16. Kelley, T. L., "The Reliability of Test Scores." Journal of Educational Research, III, (May, 1921), 374.
17. Kwalwasser-Dykema Unit E 15, descriptive material advertising Kwalwasser-Dykema Music Tests. Cooper Square: Carl Fischer, Inc., 1930, p. 1.
18. Kwalwasser, J., and G. M. Ruch, Kwalwasser-Ruch Test of Musical Accomplishment, Iowa City: published by the Extension Division, University of Iowa, 1924.
19. Kwalwasser, J., and P. W. Dykema, Kwalwasser-Dykema Tests, five Victor phonograph records, Camden: RCA Victor Co., Inc., 1930, Manual, blanks and matrices published by Carl Fischer, Inc., Cooper Square, New York.
20. Kwalwasser, J., and P. W. Dykema, Manual of Directions for Kwalwasser-Dykema Music Tests, New York City: Carl Fischer, Inc., 1930, p. 28.
21. Kwalwasser, J., Tests and Measurements in Music. Boston: C. C. Birchard & Co., 1927, p. 159.
22. Kwalwasser, J., Personal letter to H. C. LeVan, dated July 3, 1936.
23. Kwalwasser, J., "The Composition of Music Ability." National Society for the Study of Education, Yearbook, part 2, (1936), pp. 35-42.
24. Lamp, Charles J., "Can Aptitude for Specific Musical Instruments be Determined?" Music Educators National Conference, Yearbook, (1936), pp. 246-250.
25. Larson, Ruth C., "A Brief Report of a Prediction and Guidance Program in School Music." Music Educators National Conference, Yearbook, (1934), p. 224.
26. Larson, William G., "The Value of Tests and Measurements in Music Education." Music Supervisors National Conference, Yearbook, (1932), p. 324.



27. McCarty, Dorothea, "A Study of Seashore's Measures of Musical Talent." Journal of Applied Psychology, XIV, (1930), 427-455.
28. Moos, J. C., "Some Recent Developments in Music Testing." The Musical Quarterly, XIX, (July, 1933), 318-330.
29. Moos, J. C., "The Yardstick Applied to Musical Talent." The Musical Quarterly, XVI, (April, 1930), 238-262.
30. More, Grace Van Dyke, "Prognostic Testing in Music on the College Level." Journal of Educational Research, XXVI, (November, 1926), 199-212.
31. Mosher, R. M., "Improvement of Music Teaching." Music Educators National Conference, Yearbook, (1935), pp. 91-92.
32. Mursell, J. L., Human Values of Music Education. Chicago: Silver, Burdett & Co., 1934, 382 pp.
33. Mursell, J. L., and Mabelle Glenn, The Psychology of School Music Teaching. Chicago: Silver, Burdett & Co., 1931, pp. 19-20.
34. Ross, C. C., and N. T. Hooks, "How Shall We Predict High School Achievement?" Journal of Educational Research, XXII, (October, 1930), 195.
35. Rush, G. M., and G. D. Stoddard, "Criteria for the Selection of Educational Tests." Tests and Measurements in High School Instruction, Chicago: World Book Co., 1927, pp. 45-68.
36. Sanderson, Helen E., "Differences in Musical Ability of Children of Different National and Racial Origin." Journal of Genetic Psychology, XLII, (March, 1933), 100-120.
37. Seashore, C. E., Measures of Musical Talent, six Columbia phonograph records, manual and blanks, Chicago: C. H. Steeltung Co., 1919.
38. Seashore, C. E., The Psychology of Musical Talent, Chicago: Silver, Burdett & Co., 1919, 288 pp.
39. Seashore, C. E., "Psychology in Music." The Musical Quarterly, XVI, (April, 1930), 251.
40. Seashore, C. E., "The Discovery and Guidance of Musical Talent." National Society for the Study of Education, 34th Yearbook, XXI, 1935, pp. 447-461.

41. Seashore, C. E., "The Psychology of Music." Music Educators Journal, (May, 1936), p. 23.
42. Smith, Harold B., "Prognosis of Sight Singing Ability." Music Supervisors National Conference, Journal of Proceedings, (1929), pp. 437-445.
43. Stanton, "Testing the Cumulative Key for Prognosis of Musical Achievement." Journal of Educational Psychology, XXV, (1934), 45-53.
44. Stanton, H. M., and W. Koerth, "Musical Capacity Measures of Children Repeated After Musical Training." University of Iowa Studies, Series on Aims and Progress of Research, No. 42, Iowa City: University of Iowa, 1933, 1-48.
45. Sward, Keith, "Jewish Musicality in America." Journal of Applied Psychology, XVII, (May, 1933), 675-712.
46. Taylor, Cited in "Report of National Survey of Experimental Projects in Music Education." Music Educators National Conference, Yearbook, (1936), p. 278.
47. Thorndike, E. L., "The Correlation Between Interests and Abilities in College Course." Psychological Review, XXVIII, (September, 1921), 374-376.
48. Tieg, E. W., Tests and Measurements for Teachers. Boston; New York, etc., Houghton Mifflin Co., 1931, pp. 282-293.
49. Tilson, L. M., "The Music Achievement of College Students at Various Levels of Music Talent and Psychological Rating." Music Educators National Conference, Yearbook, (1935), pp. 357-363.
50. Torney, A. H., "The Concept of Validity in Mental and Achievement Testing." Journal of Educational Psychology, XXV, (February, 1934), 68.
51. Tryon, R. C., "Effect of Unreliability of Measurement on the Difference Between Groups." Journal of Comparative Psychology, VI, (1926), 449-453.
52. Watson, J. B., Behaviorism, New York: W. W. Norton & Co., Inc., 1925, p. 82.
53. Whitley, Mary T., "A Comparison of the Seashore and the Kwalwasser-Dykema Music Tests." Teachers College Record, XXXIII, (May, 1932), 731-751.
54. Wood, Ben D., Review of "The Prognostic Value of University Entrance Examinations in Scotland," Journal of Educational Research, (February, 1937), 458.

