THE EFFECTS OF FORMAL MUSIC TRAINING ON THE WING MUSICAL INTELLIGENCE SCORES

by

Jack J. Heller

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Music in the Graduate College of the State University of Iowa

August 1962

Chairman: Professor Neal E. Glenn
Co-Chairman: Professor Leonard S. Feldt

TABLE OF CONTENTS

CHAPTER		PAGE
ACKNOWLEDGMENTS		ii
LIST OF TABLES		vi
I.	SCOPE OF THE EXPERIMENT	1
	Introduction	1
	Purpose of the Study	3
	Statement of the Problem	7
	Hypotheses	7
	By-Products of the Study	8
	Need for the Study	10
	Limitations of the Study	12
II.	RELATED STUDIES	13
	Special Practice and Training Experiments	13
	Formal Training Experiments	16
III.	METHOD AND DESIGN	21
	Description of the Wing Test Battery	21
	Description of the Participating Students	23
	Selection of the Experimental Group	25
	Procedures for Testing	26
	Training of the Experimental Group	26
	Motivation for the Experimental Group	27

CHAPTER PAGE

Selection of the Control Group..... 29
Procedures for Testing..... 29
Motivation for the Control Group..... 30
Statistical Design..... 30
Equating the Groups..... 31
Size of the Groups..... 33
Analysis of Variance..... 34
Reliabilities..... 36
Intercorrelations..... 37
IV. DATA AND RESULTS..... 38
Sub-Tests 1-7..... 38
Analysis of Variance on Post-Test Scores for Sub-Tests 1-7..... 40
Sub-Tests 1-3..... 43
Analysis of Variance on Post-Test Scores for Sub-Tests 1-3..... 43
Sub-Tests 4-7..... 46
Analysis of Variance on Post-Test Scores for Sub-Tests 4-7..... 50
Reliabilities for Sub-Tests 1-7..... 52
Reliabilities for Sub-Tests 1-3..... 53
Reliabilities for Sub-Tests 4-7..... 53
Individual Score Fluctuations from Pre-Tests to Post-Tests..... 55

CHAPTER PAGE

Score Fluctuations on Sub-Tests 1-3..... 58

Score Fluctuations on Sub-Tests 4-7..... 61

Relationships between the Wing Test Battery and Scholastic Aptitudes and Achievements.., 63

V. SUMMARY AND CONCLUSIONS..... 66

Problem and Methodology..... 66

Training Effects..... 67

Stability and Reliability..... 69

Norms..... 71

APPENDIX A: OUTLINE OF MUSIC APPRECIATION COURSE.... 72

APPENDIX B: RAW SCORE DATA..... 89

BIBLIOGRAPHY..... 95

CHAPTER I

SCOPE OF THE EXPERIMENT

Introduction

Since 1901, when the first mention of the possibility of group tests of musical talent was made by Professor Carl E. Seashore,[1] a large number of experimental studies dealing with musical aptitude testing has been undertaken. The vast majority of these studies investigated various aspects of the Seashore test battery. Much valuable information has been gained from these experiments regarding the nature and measurement of musical aptitude. The need for further research remains, however, with respect to almost every aspect of this field.

Several investigators have criticized Seashore's approach to the measurement of musical aptitude.[2] Wing, for example, refers to Seashore's tests as "atomistic", and claims that tests of musical aptitude should consist

---

[1] C. E. Seashore, "Suggestions for Tests on School Children", Educ. Rev., XXII (1901), 69-82.

[2] A. W. Brown, "The Reliability and Validity of the Seashore Tests," J. Appl. Psychol., XII (1928), 468-475; C. P. Heinlein, "An Experimental Study of the Seashore Consonance Test," J. Exp. Psychol., VIII (1925), 408-432; J. L. Mursell, "Measuring Musical Ability and Achievement," J. Educ. Research, XXV (1932), 116-125.

of items that involve a musical context rather than isolated electronic tones.[3] Wing states his position as follows:

> There can be no doubt that in choosing pitch, tone, intensity, rhythm, harmony, and memory, Seashore has taken some of the commonly accepted basic qualities which, when applied to tone, are of great interest to those of high musical capacity. However, he tests for them in such an elementary form that he leaves the realm of music, which is one of the patterns and relationships in tone, and enters that of mere sensory perception --a matter outside the interest of the normal musician.[4]

After more than a decade of study, Wing published a test battery in 1948, approaching the problem of testing musical aptitude from an empirical viewpoint to determine which type of tests "prove the most efficient as judged by their agreement with music teacher's estimates."[5] After the publication of Wing's battery, numerous investigations were undertaken to examine various aspects of the tests such as reliability, validity, administrative convenience in scoring, and adequacy of the norms. These led to a revision of the battery which was published in

---

[3] H. D. Wing, "Tests of Musical Ability and Appreciation," Brit. J. Psychol., Mono. Suppl., XXVII (1948), 10.

[4] Ibid., 11.

[5] Ibid., 13.

1959.[6] At this time the test battery is regarded highly, especially in the British Isles, Australia, and New Zealand.

Purpose of the Study

The necessity for adequate tests of musical aptitude is widely recognized in this country. Before any test can be recommended, however, music educators must be convinced of its merits and must be well informed about the factors which can affect an examinee's scores. If the Wing tests are to meet the need for such an instrument, extensive investigation of its usefulness will be required.

The present study was undertaken to consider one important feature of the Wing battery -- the effects of formal music training[7] on the test scores. According to published research, this aspect of the tests has not been rigorously investigated.

Dr. Wing states:

> When pupils had continued their musical education for a year or more, the instruction received was followed by no perceptible improvement in the later test performances over and

---

[6] H. D. Wing, Standardized Tests of Musical Intelligence, National Foundation for Educational Research, London (1959).

[7] Formal music training is used to denote a normal classroom situation, as opposed to special or individual training.

above what the untrained pupils also showed, presumably as a result of general mental development during the period.[8]

Thus Wing claims that his Tests of Musical Intelligence are not unduly affected by training. In his thinking, insensitivity to training is a necessary feature of an aptitude test. One of his own criteria for any series of psychological tests of musical capacity is that "they should not be unduly influenced by training or opportunity."[9]

This statement reflects a philosophical viewpoint regarding aptitude tests which is shared by many researchers in the field of music. However, a somewhat opposing view is held by many psychologists and measurement authorities. The latter group, as typified by Wesman,[10] considers an aptitude test to be defined by its purpose, not by its content. This school of thought would regard the issue of sensitivity of the test to training to be relatively unimportant. If the test is intended for predictive purposes and serves these purposes well, then the test is employed as an aptitude measure regardless of the effects

---

[8] H. D. Wing, "A Factorial Study of Music Tests," Brit. J. Psychol., XXXI, (1941), 346.

[9] Wing, Mono. Suppl., op. cit., 55.

[10] A. G. Wesman, "Aptitude, Intelligence, and Achievement," Test Service Bulletin, Psychological Corp., New York, LI (December, 1956).

which training may have on the level of the examinee's score. Of course, if the test were very sensitive to differences in examinee background experience, it would be expected that the predictive validity would be adversely affected. Only to the extent that this actually occurs, however, would this school regard this issue of sensitivity to training to be relevant.

This study was not undertaken to support or refute Wing's concept of an appropriate test of musical aptitude, but to gather facts of importance in the interpretation of test results. The primary purpose was simply to determine whether scores on the Wing tests are sensitive to a highly specific type of training. Once this has been established, a more searching question can be examined. This deals with methods by which previous music listening skills can be quantitatively assessed and then taken into account in the norms for the test battery.

It was reported by Wing that improvement in the scores on his test battery was not appreciably greater for adolescent examinees who studied a musical instrument than for those who did not study a musical instrument.[11] Data

---

[11] Wing, Mono. Suppl., op. cit., 55.

supporting the opposite position has been presented by Steward.[12] While the effects of this type of training are important, they are probably not a crucial issue. If an individual has received training on a musical instrument, his aptitude may be more reliably inferred from his response to this training than from an aptitude test score. Thus a student who has had a moderate amount of musical instruction is not usually a candidate for aptitude testing. The influence of another type of training -- that of listening skills -- may be a more important factor in test interpretation, since almost all examinees will have had some listening experience. What is more, individual differences in listening experience are quite marked. The affect of training in listening skills has not been thoroughly investigated, yet this type of training may very well influence the test scores. Even Wing states that "listening opportunities must also be taken into account."[13]

If the extent of listening experience has an important influence on test scores, some method of taking this experience into consideration will have to be found. How-

---

[12] W. E. Steward, "An Evaluation of the Development of Appreciation for Music . . . ," (unpublished Ed.D. dissertation, Univ. of Oregon, 1960).

[13] Wing, Mono. Suppl., op. cit., 77.

ever, initial investigation of this question is best handled through a controlled training program in listening skills. While such a program does not reflect the complex state of affairs as the test administrator encounters it, the use of such a program can give a much better estimate of the training effect. A systematic approach to the teaching of listening skills was therefore employed in this study, rather than a comparison of natural groups designated as "more experienced" and "less experienced".

Statement of the Problem

The problem of this dissertation was to investigate the effects of a semester of formal music training in a music appreciation course on the scores of the Wing Standardized Tests of Musical Intelligence. The training was based on the development of listening skills and a familiarity with accepted standards of good music. A more detailed discussion of this instruction is presented in Chapter III: Training of the Experimental Group.

Hypotheses

This study tested the hypothesis that scores on the Wing Standardized Tests of Musical Intelligence would improve with a specific type of formal music training. For

statistical purposes this hypothesis was stated as a null hypothesis, and was tested at the five percent level of significance.

Two secondary hypotheses were also investigated. The first dealt with the effects of the formal training on sub-tests 1-3; the second concerned the effects on sub-tests 4-7. The first three sub-tests consist of items which measure auditory responses. It was hypothesized that these scores might be affected differently after training than the scores on the last four sub-tests, which measure musical preferences or judgements. Both of these secondary hypotheses were also tested at the five percent level of significance.

By-Products of the Study

A necessary requirement of any useful measuring instrument is that it be at least moderately reliable, that is, that it reflect systematic and lasting characteristics of the examinee. Unless a test is sufficiently reliable, it cannot provide useful information for any purpose. Since failure to reveal any significant effect of training might be misinterpreted if the tests were of low reliability, a necessary adjunct to the main analyses was the investigation of sub-test and total test reliabilities for the groups used in this study.

Another by-product of the study was the correlation of various scholastic aptitude and achievement measures of the subjects taking part in this experiment. These measures were the College Entrance Examination Board Scholastic Aptitude Test (CEEB-SAT),[14] and the Otis Intelligence Test.[15] Unstandardized measures that were included in the correlation analysis were percentile rank in high school graduating class, the cumulative grades of all college courses (based on a scale known as quality point ratio or grade point average),[16] and, finally, the grade the subjects received in the music appreciation course. These measures were correlated with the Wing test scores obtained before the training commenced. The vast majority of investigators have found a low correlation between these two types of measurements, e. g., musical aptitude and general scholastic aptitude.[17] According to several

---

[14] Scholastic Aptitude Test (New York: College Entrance Examination Board, 1960).

[15] A. S. Otis, Otis Quick Scoring Mental Ability Tests, (New York: World Book Co., 1948).

[16] This rating scale is based on the following grading scale: 40=A, 30=B, 20=C, 10=D, 0=F.

[17] R. W. Lundin, An Objective Psychology of Music (New York: The Ronald Press Co., 1953), 220-222; also, P. R. Farnsworth, The Social Psychology of Music (New York: Dryden Press, 1958), 183.

investigations by Wing, his tests correlated with intelligence scores approximately .30.[18]

Need for the Study

A reliable music aptitude test is needed in public school education in the United States. However, this evaluative instrument should not be used solely for selecting pupils for participation in musical groups. Using an evaluation tool only for such a restrictive purpose negates the basic educational philosophy of public education in this country. A reliable music aptitude test should be used for a much broader purpose -- that of improving instruction. It would be possible to investigate whether music instruction could, in fact, be aided by the use of an aptitude test. One technique by which instruction might be improved is that of grouping students with similar levels of aptitude. This procedure has not been extensively used in music education, but it has been used successfully in various other disciplines in the school curriculum.

---

[18] Wing, Mono. Suppl., op. cit., 64.