GIA Monograph Series

The Advanced Measures of Music Audiation and the Instrument Timbre Preference Test: Three Research Studies

Edwin E. Gordon

The Advanced Measures of Music Audiation and the Instrument Timbre Preference Test: Three Research Studies

by

EDWIN E. GORDON

Temple University

GIA Publications, Inc.
Chicago

TABLE OF CONTENTS

A FACTOR ANALYTIC STUDY OF THE ADVANCED MEASURES OF MUSIC AUDIATION

Introduction and Purposes

There are two important reasons for factor analyzing the 30 test questions that comprise the *Advanced Measures of Music Audiation*[1], a music aptitude test specifically designed for college and university students: 1) The test is uncommon in that through the use of a unique scoring procedure, the single test yields both a *Tonal* score and a *Rhythm* score, and, of course, a *Total* score. To add additional theoretical validity, in terms of construct validity, to what is already known about the test, it seemed important to know whether the specific test questions that are associated with the *Tonal* score account for a different factor or factors than the specific test questions that are associated with the *Rhythm* score. 2) The three option responses for every test question are "Same," "Tonal," and "Rhythm." Each test question has a musical statement and a musical answer. If the musical statement and the musical answer are the same, the correct option response is "Same." If the musical statement and the musical answer are different because there is one or more pitch changes, the correct option response is "Tonal." If the musical statement and the musical answer are different because there is one or more rhythm changes, the correct option response is "Rhythm." Considering that either a tonal or a rhythm change in a musical answer categorically constitutes a "difference" whereas no change in the musical answer constitutes "sameness," the factor analysis accords the opportunity to determine if different and same factors, in relation to the correct option responses, or if one or more mixed factors (the combining of same and different factors) are evident. Previous research using music aptitude tests with students younger than college and university age has shown that the younger the students are, the more pronounced are "difference" and "sameness" factors. Because, among other compelling reasons, the design of music aptitude tests is different for younger students who are in the developmental music aptitude stage and older students who are in the stabilized music aptitude stage, it became of interest to determine if

1. Edwin E. Gordon, *Advanced Measures of Music Audiation* (Chicago: GIA, 1989).

"sameness" and "difference" concepts consistently serve as the basis of audiation in music aptitude regardless of the chronological age of students and of their informal and formal music achievement. Moreover, it is possible that different factors may emerge because a test of developmental music aptitude is typically used with younger students and a test of stabilized music aptitude is typically used with older students.

Understanding how students conceptualize "sameness" and "difference" would seem to be of great importance in understanding how we learn when we learn music. Considering "sameness" to be a manifestation of imitation and "difference" to be dependent upon audiation, the implications for guiding preschool children through "music babble" and teaching elementary, middle and high school as well as college and university students all facets of music, particularly performance, creativity, improvisation, and reading, are enormous.

Music Aptitude and Audiation

The *Advanced Measures of Music Audiation* is a music aptitude test. It differs from the *Primary Measures of Music Audiation*[2] and the *Intermediate Measures of Music Audiation*[3] in several ways. Because the *Advanced Measures of Music Audiation* is a test of stabilized music aptitude and the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* are tests of developmental music aptitude, the designs of the two types of tests are substantially different. Both the *Primary Measures of Music Audiation* and *Intermediate Measures of Music Audiation* include two separate subtests. Each subtest, *Tonal* and *Rhythm*, is administered on different days with different verbal directions. The test questions in the *Tonal* subtest of the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* include tonal patterns without rhythm. That is, every one of the three to five pitches in a tonal pattern is of the same length. The test questions in the *Rhythm* subtest of the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* include rhythm patterns without melody. That is, every duration in all rhythm patterns is sounded on the same pitch. Another significant difference between tests of developmental and stabilized music aptitude is that because

2. Edwin E. Gordon, *Primary Measures of Music Audiation* (Chicago: GIA, 1979).
3. Edwin E. Gordon, *Intermediate Measures of Music Audiation* (Chicago: GIA, 1982).

the former is designed to be administered to younger students, the ability to read words or numerals is not required to take either the *Primary Measures of Music Audiation* or the *Intermediate Measures of Music Audiation*. All a student needs to do is circle a box in which two faces look the same if the two patterns sound the same or circle a box in which two faces look different if the two patterns sound different. "Same" and "different" are the only two option responses.

The *Advanced Measures of Music Audiation* is a self contained test with recorded directions which, nevertheless, provides for both a *Tonal* score and a *Rhythm* score. The test questions in the *Advanced Measures of Music Audiation* include both tonal and rhythm dimensions. That is, each test question includes the same short melody performed twice, or it includes two different melodies. A melody incorporates, of course, both tonal and rhythm dimensions. The *Advanced Measures of Music Audiation* is scored in a different manner from the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation*. Whereas the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* are scored in the traditional manner, one point for each question answered correctly, the *Advanced Measures of Music Audiation* is not. With regard to the *Advanced Measures of Music Audiation*, when a student believes that a musical answer is different from a musical statement because of a tonal change when, in fact, the musical statement and musical answer are the same or different because of a rhythm change, one point is subtracted from his score. Similarly, when a student believes that a musical answer is different from a musical statement because of a rhythm change when, in fact, the musical statement and musical answer are the same or different because of a tonal change, one point is subtracted from his score.

Regardless of the differences between tests of developmental and stabilized music aptitude, there is one important similarity between them. In addition to all test questions in the *Primary Measures of Music Audiation*, the *Intermediate Measures of Music Audiation*, and the *Advanced Measures of Music Audiation* being performed on a synthesizer, both types of tests are based on audiation. As a result, to be sure that a student is audiating when comparing a musical answer to a musical statement and not attempting to imitate or memorize the musical statement as a basis for comparing it to its corresponding musical answer, the amount of silent time between the end of the

musical statement and the beginning of the musical answer on the *Advanced Measures of Music Audiation* cassette recording is critical. There must be enough silent time for the student to audiate but not enough silent time for the student to imitate or memorize the musical statement. The same is true for tonal patterns and rhythm patterns on the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* cassette recordings. The first tonal pattern or rhythm pattern in a pair is considered the musical statement, and the second tonal pattern or rhythm pattern in a pair is considered the musical answer.

When audiating, the student summarizes and generalizes the musical statement. He hears only the essential parts of the melody. When imitating or memorizing, the student practices in actual time, that is, by not increasing the tempo, the musical statement from beginning to end. What is audiated contributes to a student's musicianship because it is retained. What is memorized usually does not contribute to a student's musicianship because it is quickly forgotten, particularly if no effort is made to retain it. For imitation to become useful to a student, it must serve as a prerequisite for developing skill in audiation. Considering language as a parallel, children, at most, imitate but do not memorize words to learn language. They think with words. In music, students should not memorize intervals or patterns to learn music. They audiate with patterns. Audiation is to music what thinking is to language. Simply stated, audiation is the ability to hear and to comprehend music for which the sound is no longer, or may never have been, physically present.

Previous Research

Two major studies are important to the interpretation of the results of the current study. The West Irondequoit, New York study[4] will be reviewed first and the Annville, Pennsylvania study[5] next.

The West Irondequoit, New York Study

As part of the standardization program of the *Primary Measures of Music Audiation*, 127 kindergarten children in West Irondequoit,

4. Edwin E. Gordon, *The Manifestation of Developmental Music Aptitude in the Audiation of "Same" and "Different" as Sound in Music* (Chicago: GIA, 1981).
5. Edwin E. Gordon, "The Importance of Being Able to Audiate "Same" and "Different" for Learning Music," *MEH Bulletin* (Music Education for the Handicapped), 2, 1, (Summer, 1986), pp. 3-27.

New York were administered the test battery. After relevant standardization data were derived from the answer sheets, the 40 test questions in the *Tonal* subtest were factor analyzed and the 40 test questions in the *Rhythm* subtest were factor analyzed. The principle components technique was used in conjunction with the highest multiple r to determine communality estimates. The factors were rotated to the varimax criterion of orthogonal simple structure.

Five factors were extracted from the test questions of the *Tonal* test, and seven factors were extracted from the test questions of the *Rhythm* test. Only test questions with factors loadings of .30 and higher were given consideration when interpreting a factor. No common musical characteristics were found among the test questions in a given tonal factor or among those in a given rhythm factor. What was striking was that all of the test questions in a factor had the identical correct option response. That is, the correct answer for each test question was either "same" (constituting a "same" factor) or "different" (constituting a "different" factor). Further, the bipolar factors included, of course, test questions with either a positive sign or a negative sign. For a given factor, all of the test questions with a positive sign had one identical correct option response, and all of the test questions with a negative sign had the other identical correct option response. Test questions with one sign were answered correctly, the correct answer always being "different," and test questions with the other sign were answered incorrectly, the incorrect answer always being "same."[5]

The subtests were administered again to the 111 students who remained in the class when they were in the first grade. Those results were factor analyzed. Five factors were extracted for the *Tonal*

---

5. The results of the factor analyses, although compelling, might raise some doubts. It might be said that "sameness" and "difference" factors are artifacts for one or more of the following reasons: *The order in which the test questions are presented in the subtests affects the results.* That is not the case. When the subtests were re-recorded with the test questions in random order, the results were the same. *The option responses "same" and "different" themselves create the "sameness" and "difference" factors.* The established validity of each subtest precludes that possibility. Moreover, it will be obvious from the discrimination values of the test questions reported in the test manual that the test questions are answered correctly or incorrectly according to each student's developmental music aptitudes. It will also be observed that there is no relationship between the discrimination value of a test question and the correct option response for that test question. *The students are listening in some unspecified manner to "sameness" and "difference" isolated from the music itself.* If that were the case, there would not be more than one "same" factor or more than one "different" factor in an analysis. *A "response set" produces the resultant factors.* If that were true, there would be an abundance of chance scores, and reliabilities would approach zero. The split-halves reliabilities derived from the standardization program with kindergarten children are .88 for the *Tonal* subtest, .72 for the *Rhythm* subtest, and .90 for the *Composite* test. Also, the item discrimination values range from .20 to .54. *Spurious factors, particularly difficulty factors, are disguised and appear to be "sameness" and "difference" factors.* The design of the study, including degrees of freedom and factor ratios, and the statistical analyses employed are not what may be described as the types that produce artificial "unitariness."

subtest and six factors were extracted for the *Rhythm* subtest. The bipolar factors not withstanding, only one factor associated with the *Tonal* subtest and two factors associated with the *Rhythm* subtest included only test questions that had the identical correct option response. The remaining factors were mixed. That is, they included test questions that do not always share the identical correct option response. As in the first year results, however, the factors were not representative of musical factors.

The tests were re-administered to the 87 students who had remained in the same class for the third year by the time they were in the second grade. That year, the factor analyses revealed six factors associated with the *Tonal* subtest, none bipolar, and six factors associated with the *Rhythm* subtest, only one being bipolar. Of the twelve, only two tonal factors and one rhythm factor were mixed. The structure of the factors was more like that found in the first year analysis than that found in the second. The trend was away from mixed factors and toward a return to "same" and "different" factors.

Only 82 students remained in the class when they entered the third grade. The two subtests were administered to them a fourth and final time. The factor analyses indicated a trend toward the re-establishment of mixed factors. Of the eight factors associated with the *Tonal* subtest and the seven factors associated with the *Rhythm* subtest, five of the tonal factors and six of the rhythm factors were mixed. Again, none of the factors suggested any musical characteristic. It was interesting to note that the same test questions did not constitute the "same" factors from year to year, nor did the same test questions constitute the "different" factors from year to year.

The longitudinal results of the study suggest that when students are five years old, they are preoccupied with "sameness" and "difference." When they reach eight years old, the importance of comparative "sameness" and "difference" diminishes. Nevertheless, eight year old students put emphasis on responding to "difference" to the almost exclusion of "sameness."

The Annville, Pennsylvania Study

Given the results of the previous study, the question of how factors would be constituted for students who take a stabilized music aptitude test soon after they enter the stabilized music aptitude stage

needed to be answered. Thus, the *Musical Aptitude Profile*, a stabilized music aptitude test for students in grades four through twelve, was administered to 129 students in the fourth grade in all three elementary schools in the Annville-Cleona School District in Annville, Pennsylvania.[6] Over a period of one semester, the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* were also administered to the students.

The *Musical Aptitude Profile* includes three total tests: *Tonal Imagery*, *Rhythm Imagery*, and *Musical Sensitivity*. There are seven subtests in the battery. Two subtests are provided for each of the two non-preference total tests, *Tonal Imagery* and *Rhythm Imagery*. They are *Melody* and *Harmony* for the former, and *Tempo* and *Meter* for the latter. The preference total test, *Musical Sensitivity*, consists of three subtests. They are *Phrasing*, *Balance*, and *Style*. As with the *Advanced Measures of Music Audiation*, all test questions are in the form of melodies.

When taking all subtests in the *Musical Aptitude Profile*, students are asked to compare a musical answer to a musical statement. For the *Melody* and *Harmony* subtests, students decide if the musical answer is like or different from the musical statement. For the *Tempo* and *Meter* subtests, students decide if the musical answer is the same or different from the musical answer. For the *Phrasing*, *Balance*, and *Style* subtests, students compare the musical statement with the musical answer and decide which of the two they prefer.

Even though the option responses are "like" and "different" for the *Melody* and *Harmony* subtests, for purposes of the study "like" was interpreted as if it were "same." With regard to the *Tempo* and *Meter* subtests, the option responses are already established as "same" and "different." Because the three *Musical Sensitivity* subtests are preference measures with option responses of "first" and "second" and thus could not be meaningfully translated to "same" and "different," the results of those subtests were not included in the factor analysis. In all, using the same techniques as the previous study, the results of eight subtests were factor analyzed: *Melody*, *Harmony*, *Tempo*, and *Meter* of the *Musical Aptitude Profile*, *Tonal* and *Rhythm* of the *Primary Measures of Music Audiation*; and *Tonal* and *Rhythm* of the *Intermediate Measures of Music Audiation*.

The factor analytic results of the Pennsylvania study with fourth grade students who took the *Primary Measures of Music Audiation*

6. Edwin Gordon, *Musical Aptitude Profile* (Chicago: The Riverside Publishing Company, 1988).

and the *Intermediate Measures of Music Audiation* tend to parallel those that were found with third grade students in New York who took the *Primary Measures of Music Audiation.* No "same" or "different" factors were found. All of the factors were mixed. The results for the music aptitude subtests of the *Musical Aptitude Profile*, however, were unexpected. As might be expected, the results for the *Melody*, *Tempo*, and *Meter* subtests are highly similar to those for the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation* subtests. Only mixed factors were extracted. The surprising finding was that the results for the *Harmony* subtest, the most difficult one of all in the *Musical Aptitude Profile* (a test of stabilized music aptitude) with fourth grade students closely parallel the results with kindergarten students on the *Tonal* and *Rhythm* subtests of the *Primary Measures of Music Audiation* (a test of developmental music aptitude). Specifically, the first seven factors in the analysis of the *Harmony* subtest, those which account for most of the variance, are either "same" or "different." The five remaining factors, which account for negligible variance, were mixed. None of the test questions in a given factor in all eight analyses shared similar musical characteristics.

The Present Study

A total of 5,336 undergraduate and graduate students, both music majors and non-music majors, participated in the national standardization program of the *Advanced Measures of Music Audiation.* They represented 54 institutions of higher education in 27 states in the continental United States and Canada. Of the 5,336 students, the test results of 5,250 were able to be used in the present study.

Each of the 5,250 answer sheets was scored in the following manner to conduct the factor analysis. Every test question that was answered correctly was assigned a "2." Every test question that was not answered or answered incorrectly was assigned a "1." The dichotomous data were subjected to principal axis factoring. After 283 iterations, nine factors emerged with eigenvalues of 1.00 or higher. Then a varimax rotation with Kaiser normalization was undertaken which required nine iterations.

To preserve the integrity of the answer key, the actual number of the 30 test questions as they appear on the answer sheet for the *Advanced Measures of Music Audiation* are transformed in the fol-

lowing tables. The test questions are labeled T1 through T10, R1 through R10, and S1 through S10. The 10 test question with a "T" are those which have "Tonal" as the correct answer. In those test questions, the musical answer is different from the musical statement because of one or more pitch differences. The 10 test question with an "R" are those which have "Rhythm" as the correct answer. In those test questions, the musical answer is different from the musical statement because of one or more rhythm differences. The 10 test question with an "S" are those which have "Same" as the correct answer. In those test questions, the musical statement and the musical answer are exactly the same. For purposes of interpreting the results of the factor analyses, test questions without a tonal or rhythm change in the musical answer were considered to be indicative of "sameness" and test questions with a tonal or rhythm change in the musical answer were considered to be indicative of "difference." Henceforth, the test questions will be referred to as the "T" questions, the "R" questions, and the "S" questions.

The means and standard deviations of the T, R, and S test questions are presented in Table 1. The possible mean range is from 1.0 to 2.0. Overall, the three groups of test questions have similar ranges of central tendencies and variabilities. The average of the R test questions means is slightly higher than the averages of the T and S test question means, which are highly similar. Overall, the combined T and R test questions that represent "difference" are not much more difficult than the S test questions that represent "sameness." The standard deviations of the T test questions tend to be largest and those for the S test questions tend to be smallest. Nevertheless, the standard deviations for all test questions are large in comparison to the range of the means because, as should be expected, the distributions are skewed to the left. Rarely does a student answer more test questions incorrectly than correctly.

The intercorrelations among the T test questions, among the R test questions, and among the S test questions are reported in Table 2. The correlations among the T and R test questions, among the T and S test questions, and among the R and S questions are reported in Table 3. Given the power of 5,250 subjects, the majority of the coefficients are highly significant. With regard to practical considerations, as can be seen in Table 2, the intercorrelations among the T test questions and the intercorrelations among the R test questions

are on average higher than the intercorrelations among the S test questions. The overall range of the magnitude of the coefficients is from .28 to .00. None of the coefficients is negative. More revealing, however, are the coefficients in Table 3, which range from .34 to .08. The correlations among the T and R test questions are highest, and none of the coefficients are negative. Although the correlations among the T and S test questions are higher than the correlations among the R and S test questions (both being lower than those for the T and R test questions), the two matrices include negative coefficients. The data suggest that indeed the T and R test questions have more in common with one another than either the T or R test questions have with the S test questions. Stated another way, the indication is that the T and R test questions which represent "difference" are more related to one another than to the S test questions which represent "sameness."

The communalities for the T, R, and S test questions before and after rotation may be found in Table 4. Also included in that table are the difficulty levels of the test questions, which range from .23 to .91, based upon the data obtained from this study.

The eigenvalues and percents of variance for the nine factors derived from the factor analysis before rotation and the eigenvalues and percents of variance for the nine factors derived from the factor analysis after rotation are presented in Tables 5 and 6. The nine factors account for more than 44 percent of the variance. In the unrotated analysis, the first factor accounts for approximately three times the variance of the next highest factor. In the rotated analysis, the first factor accounts for almost four times the variance of the next highest factor.

In Table 7 is the matrix of loadings for the factor analysis before rotation and in Table 8 is the matrix of loadings for the factor analysis after rotation. Except for the last three factors in the before rotation analysis, only test questions with loadings of .30 or higher are identified in the tables. It can be seen in Table 7 that of the sixteen loadings on Factor 1 before rotation, all but two are T or R test questions. S1 is dorian and S4 is atonal. Factors 2, 3, and 5 in Table 7, each with two loadings, are bipolar. One of the loadings on each factor is an S test question and the other is a T test question. Factors 4 and 6 in Table 7, each with two loadings, are mixed. Each includes a T test question and an S test question. Only one test question loads on Factors 7, 8, and 9. Each of those test questions has a loading less than .30. They are identified

because they represent clearly the highest loading on each of those factors. An S test question loads on Factor 7, an R test question loads on Factor 8, and a T test question loads on Factor 9.

Considering that all T and R test questions have musical answers that are not the same as the musical statements, the first factor in both the unrotated and rotated analyses is best interpreted as a "difference" factor. The identification of the factors becomes even clearer in the rotated analysis reported in Table 8. All nine factors are either a "difference" factor or a "same" factor. Of the nine loadings on Factor 1, all are T and R test questions, representing a "difference" factor. Similarly, Factor 2 is a "difference" factor that includes three R test questions and one T test question. Factors 3, 6, 7, and 9 include only T test questions. Factors 4, 5, and 8 include only S test questions.

It is interesting to note that the nine test questions that load on the first rotated factor have an average difficulty level of .80 and that the four test questions that load on the second rotated factor have an average difficulty level of .32. The range of the difficulty levels of the test questions that load on the remaining seven factors, each with only one loading, is from .35 to .75. That the first "difference" factor constitutes easy test questions and the second "difference" factor constitutes difficult test questions offers at least some insight into why multiple "difference" factors exist. It is compelling that the major portion of the variance of the rotated factor analysis is accounted for by two "difference" factors of vastly different difficulty levels and not even one "same" factor. With regard to the construct validity of the *Advanced Measures of Music Audiation*, however, the data do not substantiate distinct tonal and rhythm factors.

Table 1
Means and Standard Deviations of the T, R, and S Test Questions Constituting the Advanced Measures of Music Audiation

Item	Mean	SD
T1	1.42	.4935
T2	1.62	.4853
T3	1.48	.4995
T4	1.85	.3530
T5	1.71	.4546
T6	1.51	.5000
T7	1.81	.3941
T8	1.58	.4941
T9	1.53	.4993
T10	1.28	.4474
R1	1.91	.2866
R2	1.36	.4794
R3	1.64	.4813
R4	1.94	.2427
R5	1.75	.4341
R6	1.88	.3246
R7	1.85	.3568
R8	1.28	.4474
R9	1.29	.4531
R10	1.36	.4794
S1	1.55	.4974
S2	1.61	.4877
S3	1.67	.4717
S4	1.75	.4325
S5	1.35	.4763
S6	1.85	.3525
S7	1.39	.4887
S8	1.23	.4208
S9	1.64	.4807
S10	1.56	.4960

Table 2
Intercorrelations Among the T Test Questions, the R Test Questions, and the S Test Questions Constituting the Advanced Measures of Music Audiation

T Test Questions

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
T1	—	.13	.15	.17	.19	.10	.20	.22	.10	.14
T2	.13	—	.07	.16	.20	.08	.16	.16	.14	.09
T3	.15	.07	—	.14	.14	.10	.16	.19	.10	.14
T4	.17	.16	.14	—	.17	.07	.19	.23	.13	.07
T5	.19	.20	.14	.17	—	.08	.27	.26	.16	.11
T6	.10	.08	.10	.07	.08	—	.06	.08	.08	.03
T7	.20	.16	.16	.19	.27	.06	—	.28	.14	.09
T8	.22	.16	.19	.23	.26	.08	.28	—	.13	.15
T9	.10	.14	.10	.13	.16	.08	.14	.13	—	.06
T10	.14	.09	.14	.07	.11	.03	.09	.15	.06	—

R Test Questions

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
R1	—	.02	.13	.22	.17	.24	.19	.02	.00	.04
R2	.02	—	.08	.05	.08	.06	.06	.12	.05	.14
R3	.13	.08	—	.08	.12	.18	.14	.08	.00	.08
R4	.22	.05	.08	—	.15	.25	.19	.02	.02	.05
R5	.17	.08	.12	.15	—	.19	.17	.07	.04	.08
R6	.24	.06	.18	.25	.19	—	.23	.05	.02	.06
R7	.19	.06	.14	.19	.17	.23	—	.02	.01	.07
R8	.02	.12	.08	.02	.07	.05	.02	—	.00	.13
R9	.00	.05	.00	.02	.04	.02	.01	.00	—	.00
R10	.04	.14	.08	.05	.08	.06	.07	.13	.00	—

S Test Questions

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	—	.00	.08	.15	.00	.06	.08	.10	.07	.02
S2	.00	—	.00	.03	.06	.02	.04	.02	.03	.01
S3	.08	.00	—	.13	.04	.08	.09	.07	.00	.07
S4	.15	.03	.13	—	.00	.10	.07	.07	.14	.08
S5	.00	.06	.04	.00	—	.04	.04	.05	.00	.00
S6	.06	.02	.08	.10	.04	—	.01	.03	.08	.03
S7	.08	.04	.09	.07	.04	.01	—	.01	.08	.04
S8	.10	.02	.07	.07	.05	.03	.01	—	.01	.09
S9	.07	.03	.00	.14	.00	.08	.08	.01	—	.01
S10	.02	.01	.07	.08	.00	.03	.04	.09	.01	—

Table 3
Correlations Among the T and R Test Questions, the T and S Test Questions, and the R and S Test Questions Constituting the Advanced Measures of Music Audiation

T and R Test Questions

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
R1	.12	.14	.15	.07	.10	.15	.09	.05	.01	.08
R2	.14	.14	.12	.11	.11	.16	.11	.08	.03	.11
R3	.08	.07	.21	.06	.10	.10	.08	.11	.01	.12
R4	.19	.09	.17	.29	.23	.29	.21	.04	.00	.06
R5	.17	.12	.18	.17	.17	.23	.19	.07	.01	.07
R6	.08	.06	.08	.07	.08	.09	.07	.05	.00	.04
R7	.19	.09	.18	.19	.23	.34	.23	.06	.02	.07
R8	.15	.15	.09	.11	.15	.22	.18	.09	.02	.09
R9	.08	.11	.11	.09	.11	.12	.09	.09	.07	.09
R10	.12	.13	.08	.00	.02	.05	.03	.14	.01	.12

T and S Test Questions

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
S1	27	-.03	.07	.14	-.08	.07	.04	.04	.07	.02
S2	.11	.11	.08	.10	-.06	.03	.03	.05	.03	.02
S3	.12	.03	.11	.09	-.04	.06	.04	.07	.09	.03
S4	.13	.00	.11	.23	-.07	.11	.10	.03	.13	.02
S5	.13	-.01	.13	.25	-.07	.11	.05	.03	.12	.02
S6	.02	-.01	.03	.16	-.02	.00	.07	.05	.00	.04
S7	.13	-.01	.12	.20	-.02	.11	.09	.02	.15	.03
S8	.18	-.03	.14	.20	.01	.13	.09	.10	.13	.09
S9	.06	.00	.04	.07	-.03	.04	.08	.17	.03	.02
S10	.07	-.02	.06	.07	-.02	.01	.02	.14	.01	.04

R and S Test Questions

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
S1	.08	.00	.10	.16	-.02	.09	.07	.02	.08	.01
S2	.12	.03	.04	.06	-.02	.02	.03	.09	.05	.04
S3	.11	-.01	.10	.11	-.05	.06	.05	.05	.09	.01
S4	.08	.03	.10	.16	-.01	.13	.06	.03	-.02	.05
S5	.11	-.01	.06	.12	-.02	.11	.09	.02	.10	.02
S6	.12	.00	.13	.19	-.04	15	.09	.01	.13	.03
S7	.10	.00	.08	.16	-.05	.12	.11	.05	.12	.04
S8	.05	.00	.02	.00	-.04	-.01	.02	.08	.06	.01
S9	.00	-.03	.00	-.02	.00	.00	.00	.01	.02	.07
S10	.10	.01	.04	.07	.00	.03	.00	.10	.03	.06

Table 4
Communalities and Difficulty Levels for the T, R, and S Test Questions Constituting the Advanced Measures of Music Audiation

Test Question	Difficulty Index	Communality After Rotation	Communality Before Rotation
T1	.42	.16604	.23999
T2	.62	.13025	.52959
T3	.48	.12940	.33356
T4	.85	.23844	.28558
T5	.71	.20822	.25301
T6	.51	.05800	.07492
T7	.81	.26110	.33801
T8	.58	.21598	.31224
T9	.53	.10651	.81659
T10	.28	.09203	.19276
R1	.91	.14646	.18961
R2	.36	.10318	.15000
R3	.64	.12399	.17284
R4	.94	.16369	.28916
R5	.75	.11677	.14133
R6	.88	.23770	.32464
R7	.85	.14847	.19116
R8	.28	.06365	.10680
R9	.29	.02312	.01856
R10	.36	.07425	.14445
S1	.55	.12985	.60142
S2	.61	.03858	.13258
S3	.67	.07321	.11007
S4	.75	.16870	.84200
S5	.35	.04177	.09898
S6	.85	.06294	.07031
S7	.39	.04803	.08458
S8	.23	.08413	.16488
S9	.64	.07223	.09774
S10	.56	.03494	.05804

Table 5
Principal Axis Factoring for the Test Questions Constituting the Advanced Measures of Music Audiation

Factor	Eigenvalue	Percent of Variance	Cumulative Percent
1	4.03087	13.4	13.4
2	1.59122	5.3	18.7
3	1.26202	4.2	22.9
4	1.15231	3.9	26.8
5	1.10014	3.7	30.5
6	1.04793	3.4	33.9
7	1.04535	3.5	37.4
8	1.02130	3.4	40.8
9	1.00866	3.4	44.2

Table 6
Varimax Rotation With Kaiser Normalization for the Test Questions Constituting the Advanced Measures of Music Audiation

Factor	Eigenvalue	Percent of Variance	Cumulative Percent
1	3.35188	11.2	11.2
2	.90026	3.0	14.2
3	.72117	2.4	16.6
4	.64426	2.1	18.7
5	.49628	1.7	20.4
6	.44671	1.5	21.9
7	.36696	1.2	23.1
8	.22374	.7	23.8
9	.21370	.7	24.5

Table 7
Unrotated Factor Matrix for the Test Questions Constituting the Advanced Measures of Music Audiation

Test Questions	Factors 1	2	3	4	5	6	7	8	9
T1	.385								
S1	.371		.490		-.393	.332			
R1	.367								
T2	.379				.429	.396			
T3	.345								.286
R3	.362								
R4	.377							.216	
T4	.498								
T5	.476								
S4	.498	-.350		.627					
R5	.350								
T7	.534								
R6	.499								
S5							.250		
T8	.497								
R7	.393								
T9	.357	.567	-.455	.336					

Table 8
Rotated Factor Matrix for the Test Questions Constituting the Advanced Measures of Music Audiation

	Factors								
Test Questions	1	2	3	4	5	6	7	8	9
S1					.738				
R1	.428								
R2		.325							
T2						.620			
T3							.476		
R4	.490								
T4	.507								
T5	.375								
S4				.852					
R5	.349								
T7	.524								
R6	.560								
S5								.300	
T8	.355								.302
R5	.427								
T9			.872						
T10		.403							
R8		.315							
R10		.356							

Discussion and Conclusions

In previous research it was found that kindergarten children attend to "difference" and "sameness" in music. As they get older, they gradually become less concerned with the distinction between "difference" and "sameness." Nonetheless, to some extent older children still attend to "difference" and "sameness" in music, depending upon to what they are listening. Such proclivities and change may be attributed to the fact that kindergarten children are in the developmental music aptitude stage and older children are in the stabilized music aptitude stage. More convincing, however, may be whether a developmental or stabilized music aptitude test is administered to the children. The content and constructs of the two types of tests are different. The results of the present study indicate that adults to whom, of course, a stabilized music aptitude test is administered, are primarily concerned with "difference" and "sameness."

The implications of the ongoing research for music education are provocative. Regardless of their chronological age, level of music aptitude, level of music achievement, or whether they are receiving general music, choral music, or instrumental music instruction, all students, especially the very youngest, should be given opportunities to engage in creativity and improvisation activities. Because imitation serves as a readiness for audiation, imitation must command primary attention. Just as soon as students are capable in imitation, however, they should be engaged in audiation activities. Imitation is associated with "sameness," whereas audiation is associated with "difference." Unless students are given specific guidance in audiating differences in music, in terms of discrimination and inference types of learning, they will not adequately achieve, in terms of comprehension, in reproducing the music of others. To insist that the memorization of music or the ability to read notation for performance purposes are principal goals of music education is becoming more and more questionable.

Bibliography

Clark, E. V. (1973). *What's in a Word? On the Child's Acquisition of Semantics in His First Language.* New York: Academic Press.

Bessemer, D. W. (1975). *Knowledge of the Meaning of the Terms "Same" and "Different.* Los Alamitos, California: SWRL (Southwest Regional Laboratory).

Bessemer, D. W. and M. J. Fischer. (1973). *Performance of Equivalence Discrimination Tasks by Kindergarten Children.* Los Alamitos, California: SWRL (Southwest Regional Laboratory).

Goodman, N. (1972). *Problems and Projects.* Indianapolis, Indiana: Bobbs-Merrill.

Gordon, E. E. (1979). *Primary Measures of Music Audiation.* Chicago: GIA.

Gordon, E. E. (1981). *The Manifestation of Developmental Music Aptitude in the Audiation of "Same" and "Different" as Sound in Music.* Chicago: GIA.

Gordon, E. E. (1982). *Intermediate Measures of Music Audiation.* Chicago: GIA.

Gordon, E. E. (1986). "The Importance of Being Able to Audiate "Same" and "Different" for Learning Music." *MEH Bulletin* (Music Education for the Handicapped), 2, 1, 3-27.

Gordon, E. E. (1986). *The Nature, Description, Measurement, and Evaluation of Music Aptitudes.* Chicago: GIA.

Gordon, E. (1988). *Musical Aptitude Profile.* Chicago: The Riverside Publishing Company.

Gordon, E. E. (1989). *Advanced Measures of Music Audiation.* Chicago: GIA.

Webb, R. A., Oliveri, M. E., and L. O'Keefe. (1974). "Investigations of the Meanings of "Different" in the Language of Young Children." *Child Development,* 45, 984-991.

CONTINUING STUDIES OF THE CHARACTERISTICS OF THE INSTRUMENT TIMBRE PREFERENCE TEST*

Introduction

The *Instrument Timbre Preference Test*[1] includes seven different synthesized timbres of the same brief melody which are performed with the same musical expression on a Moog Opus 3 Synthesizer. The different timbres are produced by changing and combining footages (octaves) and by modifying tone color through filtering. Each of the derived seven synthesized timbres is intended to represent the actual timbre of one or more woodwind or brass instruments. The first represents the flute; the second, clarinet; the third, saxophone and French horn; the fourth, oboe, English horn, and bassoon; the fifth, trumpet and cornet; the sixth, trombone, baritone, and French horn; and the seventh, tuba and Sousaphone. Because the melody has a range of an eleventh, it is possible to produce a broad spectrum of each timbre.

The purpose of the *Instrument Timbre Preference Test* is to act as an objective aid to the teacher and parent in helping a student choose an appropriate woodwind or brass instrument to learn to play in beginning instrumental music. Timbre preference is perhaps next in importance to music aptitude in determining the extent to which a student will be successful in learning to play an instrument. Results on the test, for which there are no correct or incorrect answers or norms, indicate a student's likes and dislikes of sounds associated with various woodwind and brass instruments. If a student likes the sound associated with a particular instrument, it is likely that he or she will be more successful on that instrument than on an instrument which has a sound that he or she does not like or may even dislike.

There are various reasons for using synthesized timbres rather than actual timbres on the *Instrument Timbre Preference Test*. It was found in research in the development of the test that primary among those reasons is that although a student may claim to prefer a timbre that is heard on an actual instrument, he or she may choose that timbre simply because of familiarity with the instrument itself. Or, the

* Another version of this study may be found in the *Council For Research In Music Education*, Fall, 110, 1991.

1. Edwin E. Gordon, *Instrument Timbre Preference Test* (Chicago: GIA, 1984).

student may reject a timbre because of previous unfortunate experiences. Perhaps most compelling is that a student may choose a timbre that is heard performed on an actual instrument because he or she wants to play that instrument for reasons other than its timbre. The student may associate the instrument with male or female performers, or the student may know that if he or she plays that instrument, it will be possible to become a member of a certain music organization and participate in extra-musical activities. Students may choose a timbre heard performed on an actual instrument because their parents wish them to play that instrument, because one of their friends plays or wants to play that instrument, because a famous artist plays that instrument, because they associate that instrument with a favorite piece of music, or because an adult insists on an instrument that can be easily transported.

The *Instrument Timbre Preference Test* includes approximately 22 minutes of listening time, including the recorded directions. The seven timbres are organized into 42 recorded test items on the cassette. Each of the seven timbres is paired twice with every other timbre, allowing each timbre to be heard twelve times. It is heard six times first in the pair and six times second in the pair. Each pair of timbres is a test item. The student is asked to listen to each test item and to indicate on an answer sheet which one of the two timbres he or she prefers. Because the melody is the same, is in the same key, and musical expression is held constant for every timbre, timbre is the only factor that changes in the test item. The melody does not become boring to the students, because it is uncommon, ambiguous, and brief.

Background of Research

Considerable developmental research, specifically with regard to the reliability and construct validity of the *Instrument Timbre Preference Test*, is reported in the test manual.[2] Since the publication of the test, however, two important research projects have been completed. They are practical studies of the longitudinal predictive validity of the *Instrument Timbre Preference Test*.[3] Those two

2. Edwin E. Gordon, Part Six of the *Manual for the Instrument Timbre Preference Test* (Chicago: GIA, 1984), pp. 22-49.

3. Edwin E. Gordon, "Final Results of a Two-Year Longitudinal Predictive Validity Study of the Instrument Timbre Preference Test and the Musical Aptitude Profile," *Council for Research in Music Education*, 89, (Fall, 1986), pp. 8-17 and Edwin E. Gordon, *A Two-Year Longitudinal Predictive Validity Study of the Instrument Timbre Preference Test and the Intermediate Measures of Music Audiation* - Predictive Validity Studies of IMMA and ITPT, GIA Monograph Series - (Chicago: GIA, 1989).

studies are summarized below. The reader may also be interested in studies which bear on various philosophical aspects and on the theoretical validity of the *Instrument Timbre Preference Test* which have been conducted by other investigators.[4]

The first predictive validity study was conducted in all fifth grades in the Rush-Henrietta Central School District in Henrietta, New York. The *Musical Aptitude Profile* and the *Instrument Timbre Preference Test* were administered to all students. In the first year of the study, 57 students (the experimental group) in beginning instrumental music classes elected to learn to play instruments for which they had demonstrated a timbre preference on the *Instrument Timbre Preference Test*. One-hundred eleven students (the control group) elected to learn to play instruments that were not suggested by their scores on the *Instrument Timbre Preference Test*. The students received one 30 minute group homogeneous instrumental lesson each week. Experimental and control group students were taught together in one class. At the end of each academic year, the students read and performed three short etudes on their instruments. One etude was prepared with teacher help, another was prepared without teacher help, and the third was sightread. The recorded performances were evaluated independently by two judges.

After the first year of instruction, the students in the experimental group scored significantly higher on the achievement criteria than did students in the control group. Moreover, in the experimental group, 24 (48 percent) of the students discontinued instruction, whereas 64 (58 percent) of the students in the control group discontinued instruction. The predictive validity of the *Musical Aptitude Profile* when used in conjunction with the *Instrument Timbre Preference Test* was found to be .81. After the second year of instruction, the students in the experimental group again scored significantly higher on the achievement criteria than did students in the control group. Also, 22 students (39 percent) in the experimental group and 28 students (25 percent) completed two years of instruction. The predictive validity of the two tests combined was found to be even higher in the second year; the coefficient was .85. The data indicate that approximately 72 percent of the reason for students' success or

---

4. Charles P. Schmidt and Barbara E. Lewis, "A Validation Study of the Instrument Timbre Preference Test, *"Psychology of Music*, 16, 2, (1988), pp. 143-155; Don D. Coffman and Karin Harfst Sehmann, "Musical Instrument Preference," *Update*, 7, 2, (1989), pp. 32-34; and Robert A. Cutietta and Maria Foustalieraki, "Preferences for Select Band and Non-Band Instrument Timbres Among Students in the United States and Greece," *Council for Research in Music Education*, 105, (Summer, 1990), pp. 72-80.

lack of success in beginning instrumental music is a result of a combination of their music aptitude and favorable circumstances which encourage and allow them to play a music instrument for which they have a timbre preference. That is, when used in conjunction with the *Musical Aptitude Profile*, the *Instrument Timbre Preference Test* increases accuracy of prediction of success in beginning instrumental music by approximately 16 percent.

The second predictive validity study, designed much like the first one, was conducted in all fourth grades in the Guilderland School District in Guilderland, New York. The *Intermediate Measures of Music Audiation* and the *Instrument Timbre Preference Test* were administered to all students. In the first year of the study, 30 students (the experimental group) in beginning instrumental music classes elected to learn to play instruments for which they had demonstrated a timbre preference on the *Instrument Timbre Preference Test*. One-hundred fifty-one students (the control group) elected to learn to play instruments that were not suggested by their scores on the *Instrument Timbre Preference Test*. The students received at least one group homogeneous instrumental lesson each week. Experimental and control group students were taught together in one class. At the end of the first academic year, the students read and performed two short etudes on their instruments. One was prepared with teacher help and the other without teacher help. At the end of the second year, the students also sightread an etude. The recorded performances were evaluated independently by the same two judges who participated in the first study.

After the first year of instruction, the students in the experimental group scored significantly higher on the achievement criteria than did students in the control group. At the end of the first semester, 49 students (32 percent) in the control group discontinued instruction, and 2 (6 percent) in the experimental group discontinued instruction. At the end of the second semester, 39 of the 102 students (38 percent) in the control group who continued instruction after the first semester discontinued instruction, and 9 of the 28 students (32 percent) in the experimental group who continued instruction after the first semester discontinued instruction. The predictive validity of the *Intermediate Measures of Music Audiation* when used in conjunction with the *Instrument Timbre Preference Test* was found to be .48. After the second year of instruction, the students in the experimental group

again scored significantly higher on the achievement criteria than did students in the control group. The discontinuance of instruction for the first and second semesters reflected the ratios of the first and second semesters of the first year; relatively more students in the experimental group continued instruction. The predictive validity of the two tests combined was .80 at the end of the second year, higher than when the *Intermediate Measures of Music Audiation* is used alone. The data indicate that approximately 64 percent of the reason for students' success or lack of success in beginning instrumental music is a result of a combination of their music aptitude and timbre preference. It was concluded that students' high music aptitude will not compensate for their learning to play an instrument for which they do not have a timbre preference nor will students' timbre preference for a music instrument compensate for their low music aptitude in learning to play that instrument.

Purpose and Problems of the Study

The practical aspect of any test, particularly its predictive validity, is of central importance to educators. That, however, should not preclude the acquisition of a better theoretical understanding of the characteristics of a unique test like the *Instrument Timbre Preference Test*. The supplementary information that the test offers could indirectly affect related research and directly affect music instruction. Perhaps even more important, relevant research could offer direction for the improvement of the test itself. With those possibilities in mind, the present study was undertaken. The following are the more important questions that were addressed in conjunction with a representative sample of typical elementary school students who were eligible to participate and those who actually participated in a beginning instrumental music program.

1) Do boys and girls have different timbre preferences?
2) Do students who have more than one instrument timbre preference have a stronger preference for any one timbre than do students who have only one timbre preference?
3) Are certain timbres or grouping of timbres preferred more than others?

4) Does the preference for one timbre or groupings of timbres indicate the preference or non-preference of other timbres or groupings of timbres?
5) How many students have one or more timbre preferences?
6) How many students have no timbre preference?
7) How many students dislike certain timbres?
8) Are there differences in the pattern of timbre preferences between students who are considering participation in a beginning instrumental music program and students who actually enter the program?
9) Do students who have high music aptitude have stronger timbre preferences than do students who have low music aptitude?
10) Do students who have high music aptitude have different patterns of timbre preferences from students who have low music aptitude?
11) Do teachers evaluate differently the music achievement of students who do not share the same timbre preferences?
12) Do students who study an instrument for which they have a timbre preference demonstrate more success in beginning instrumental music than students who study an instrument for which they do not have a timbre preference?

Design of the Study

The two-year study took place from 1987 through 1989 in the North Babylon Union Free School District in North Babylon, New York.[5] The *Instrument Timbre Preference Test* and the *Intermediate Measures of Music Audiation* were administered to all 258 students in the fourth grade who were enrolled in the five elementary schools in the school district. Of the 258 students, 78 elected to begin the study of a music instrument in September, 1987. They participated in at least one group lesson each week.

Although the teachers had knowledge of the students' test scores and encouraged those with high overall music aptitude to enroll in beginning instrumental music without discouraging those with lower overall music aptitude from participating in the program, no effort was made to direct students to choose instruments to study that were in accordance with their timbre preferences. The teachers, however,

5. The writer is grateful to James R. Brown, Director of Music, and to the teachers in the Department of Music. Without their interest, cooperation, and assistance, the study could not have taken place.

made the students aware of their timbre preference scores but not of their music aptitude scores. The students were guided in choosing an instrument to study according to traditional practices. In the event, however, that a student chose to study an instrument that was suggested by *Instrument Timbre Preference Test* results but he or she did not have personal access to one and could not afford to rent or buy one, the teachers were able to offer the student the use of an appropriate instrument free of charge.[6]

In May, 1989, at the conclusion of two years of instrumental music instruction, 63 of the 78 students who initially entered the program remained. Ten students were playing a woodwind or brass instrument that was suggested by their *Instrument Timbre Preference Test* results and 47 were not. Six students were playing drums. There were, of course, additional students who entered the instrumental music program during the first year and at the beginning of and during the second year of instruction. They were not included in the study. Because many of them were new to the school district, test data were not available for the majority of late registrants.

At the conclusion of two years of instrumental music instruction, the five teachers who taught the students in the five elementary schools were asked to rate the 57 students' instrumental performance achievement. That was accomplished without the students performing any etudes. Each teacher did the ratings of his or her students independently of the other teachers on the basis of assessing how well each student had performed in lessons and in band activities over the previous two years. The teachers did arrive at a consensus beforehand for awarding the ratings. Nonetheless, though an impractical situation, it may be assumed that the validity of the ratings (which is unknown) would have been higher if only one teacher taught all of the students and completed all of the ratings. The rating scale may be found at the end of this report. Each student was given a 1 through 5 for tonal achievement, a 1 through 5 for rhythm achievement, and a 1 through 5 for overall achievement.[7]

Means, standard deviations, correlations, and percentages pertaining to *Instrument Timbre Performance Test*, *Intermediate Measures of Music Audiation*, and rating scale scores were com-

6. The instruments that were lent to the students were those that were granted by the National Association of Band Instrument Manufacturers (NABIM) for the purpose of continuing research in timbre preference. The same instruments had been used by the students in the Rush-Henrietta and Guilderland schools.

7. Because teachers are accustomed to giving a 1 for the highest level of achievement, they were given such directions on the rating scale form. For purposes of statistical analyses, however, 1 was converted to 5, 2 to 4, 4 to 2, and 5 to 1. A 3, of course, remained the same.

puted and compared for the 258 students in the entire fourth grade and for only those 57 students who completed two years of instrumental music instruction. The following results are based upon those analyses.

Results of the Study

The means and standard deviations of the seven scores on the *Instrument Timbre Preference Test* for all 258 students and for only the 57 students who completed two years of instrumental instruction are reported in Table 1. The data were collected at the beginning of the year when all students were entering the fourth grade. In all tables, Timbre A represents the flute; Timbre B, clarinet; Timbre C, saxophone and French horn; Timbre D, oboe, English horn, and bassoon; Timbre E, trumpet and cornet; Timbre F, trombone, baritone, and French horn; and Timbre G, tuba and Sousaphone.

It can be seen in Table 1 that the relationships and magnitudes among the means and standard deviations for the instrumental group is essentially a microcosm of the means and standard deviation for the larger group. That would suggest that the overall timbre preferences of students who enroll in beginning instrumental music are highly similar to those of students at large. It is also obvious in Table 1 that, in general, timbres associated with woodwind instruments are more attractive to students than are timbres associated with brass instruments. The means for Timbre E, although associated with brass instruments, are comparable to those of Timbres B and C which are associated with woodwind instruments. Thus it might seem that students prefer higher rather than lower sounds, regardless of timbre. That, however, is not necessarily borne out in the intercorrelations among the students' timbre preferences.

In Table 2 are the intercorrelations among the timbre preferences for all 258 fourth grade students. The data indicate, for example, that although students who prefer Timbre A (flute) tend to have opposite preferences for Timbres F and G (lower brasses), they do not demonstrate strong, if any, timbre preferences for other woodwind instruments that also have higher ranges. Moreover, students with a preference for Timbre G (tuba and Sousaphone) show only a moderate preference for Timbre F (trombone and baritone), all of which are in lower ranges. Simply, there appears to be no clear pattern of related timbre preferences in terms of instrument type or

range. From the point of view of construct validity, predictive validity notwithstanding, that finding in addition to the number and magnitude of negative intercorrelations and the moderate to low positive intercorrelations is an indication that the psychological constructs underlying the design of the test have merit.

Table 1
Instrument Timbre Preference Test Means and Standard Deviations for All Students in the Fourth Grade and for Those Who Completed Two Years of Instrumental Music Instruction

	Entire Fourth Grade		Students Who Completed Two Years of Instrumental Music Instruction	
	Mean	SD	Mean	SD
Timbre				
A	8.1	2.74	7.9	2.79
B	6.5	2.09	6.3	1.95
C	6.5	2.33	6.5	2.62
D	7.9	2.52	7.7	2.93
E	5.9	2.14	6.2	2.31
F	4.4	2.04	4.4	2.21
G	2.7	2.89	3.1	3.45

Table 2
Intercorrelations Among the Timbres of the Instrument Timbre Preference Test Based on the Results of All Fourth Grade Students

	Timbre						
	A	**B**	**C**	**D**	**E**	**F**	**G**
Timbre							
A		-.10	-.19	.29	.09	-.59	-.63
B	-.10		.30	-.33	-.35	-.12	-.23
C	-.19	.30		-.34	-.57	-.16	.00
D	.29	-.33	-.34		.06	-.39	-.41
E	.09	-.35	-.57	.06		.02	-.19
F	-.59	-.12	-.16	-.39	.02		.40
G	-.63	-.23	.00	-.41	-.19	.40	

As can be seen in Table 3, there are no important relationships between sex and instrument timbre preference in terms of type or range of instruments. That is, the bi-serial coefficients (2=girls, 1=boys) indicate that there is no more than four percent in common between sex and any instrument timbre preference. In research using actual instruments, both with and without sound, that is not always the case. It would seem that either the synthesized timbres on the *Instrument Timbre Preference Test* are not truly representative of the actual timbres of the instruments they are intended to represent or the test is indeed successful in preventing stereotypes and other extra-musical factors from influencing students in choosing instruments that may be musically inappropriate for them to learn to play.

In accordance with the guidelines in the *Instrument Timbre Preference Test* Manual, students are designated as having a preference for a timbre if they select that timbre ten or more of the twelve times it is heard in the test. If students select a timbre two or fewer of the twelve times it is heard in the test, they are designated as having a dislike for that timbre. Students who select a timbre no fewer than three times or no more than nine of the twelve times that it is heard in the test are designated as being indifferent to that timbre. If the majority of the students who take the test were designated as being

indifferent to all seven timbres, that is, if they did not prefer any timbre or if they did not dislike any timbre, the value of administering the test would be called into question. It was found, however, that of the 258 fourth grade students who took the test, only 49 (19 percent) were indifferent to all seven timbres; 122 (47 percent) had a preference for one timbre; 85 (33 percent) had a preference for two timbres; and 2 (1 percent) had a preference for three timbres. Because the majority of students, more than 80 percent, have at least one timbre preference, the importance of a teacher using such objective information in assisting students in choosing specific instruments to study cannot be overestimated.

Table 3
Correlations Between Sex and Instrument Timbre Preference Test Scores Based on the Results of All Fourth Grade Students

Timbre	
A	.17
B	-.02
C	-.17
D	.10
E	.18
F	-.19
G	.00

Mean scores indicate that groups of students have a greater preference for woodwind timbres than for brass timbres. The extent to which students show a preference for a given timbre by choosing it ten or more of the times that it is heard in the test or show a dislike of a given timbre by choosing two or fewer of the times that it heard in the test offers greater specificity to the matter. Relevant data are presented in Table 4 for the all fourth grade students and in Table 5 for only students who completed two years of instrumental music instruction. The trend of percentages in the two groups is quite similar. Somewhat less than fifty percent of students prefer Timbre A, that for flute, and Timbre D, that for oboe, English horn, and bassoon, whereas somewhat more than fifty percent of students dislike Timbre G, that for Sousaphone and tuba.

All students who have only one timbre preference do not, of course, share the same timbre preference. Similarly, it was found that students who have two timbre preferences do not strongly prefer or dislike any one timbre or any consistent group of timbres. The coefficients are reported in Table 6. The data suggest the slight possibility, however, that students who indicate a preference for two timbres tend to be unfavorably inclined toward Timbre B, that for clarinet. With regard to woodwind and brass timbres in general, 20 (8 percent) of all 258 students had preferences for at least one woodwind instrument and one brass instrument and 6 (10 percent) of the 57 students who completed two years of instrumental instruction had preferences for at least one woodwind instrument and one brass instrument.

Table 4
Number and Percent of All Fourth Grade Students Who Indicated a Preference for and a Dislike of Each Timbre on the Instrument Timbre Preference Test

	Chosen 10 to 12 Times		Chosen 0 to 2 Times	
	Number	Percent	Number	Percent
Timbre				
A	97	38	11	4
B	28	11	9	3
C	36	14	8	3
D	80	31	6	2
E	33	13	14	5
F	12	5	45	17
G	18	7	155	60

Table 5
Number and Percent of Students Who Completed Two Years of Instrumental Music Instruction Who Indicated a Preference for and a Dislike of Each Timbre on the Instrument Timbre Preference Test

	Chosen 10 to 12 Times		Chosen 0 to 2 Times	
	Number	Percent	Number	Percent
Timbre				
A	20	35	4	7
B	3	5	3	5
C	8	14	4	7
D	19	33	4	7
E	9	16	4	7
F	4	7	11	19
G	7	12	33	58

The music aptitude scores of the 258 students in the fourth grade were only slightly lower, approximately one point, and a little more homogeneous than the scores of typical fourth grade students in the United States.[8] Means and standard deviations for the 258 students on the *Intermediate Measures of Music Audiation* may be found in Table 7. The intercorrelation between the students' scores on the *Tonal* and *Rhythm* subtests of the battery was .38. It is .40 for students at large. The scores were used to make comparisons between the music aptitudes of the 258 students and their timbre preferences.

Some persons believe that students who have high levels of music aptitude have timbre preferences and students who have low levels of music aptitude have no timbre preferences or, in fact, a dislike of any timbre. That is not the case. Correlations between scores on the *Intermediate Measures of Music Audiation* and the number of times each timbre was preferred on the *Instrument Timbre Preference Test* by all 258 fourth grade students are presented in Table 8. The highest correlation (.12), that for Timbre B with overall music aptitude, suggests that there is less than one and a half percent in common between music aptitude and specific instrument timbre preferences. Moreover, level of music aptitude is in no way related to overall instrument timbre preference. When comparative levels of music

8. For comparative standardization data, See Part Seven of the Manual for the *Primary Measures of Music Audiation and the Intermediate Measures of Music Audiation* by Edwin E. Gordon. (Chicago: GIA, 1986).

Table 6
Correlations Between Instrument Timbre Preferences Test Scores and Number of Preferences Based on the Results of All Fourth Grade Students

Timbre	
A	.05
B	-.23
C	-.13
D	.06
E	.09
F	-.01
G	.06

Table 7
Intermediate Measures of Music Audiation Means and Standard Deviations for All Students in the Fourth Grade

	Mean	SD
Tonal	34.2	2.58
Rhythm	32.4	3.17
Composite	66.6	4.70

Table 8
Correlations Between Scores of All Fourth Grade Students on the Instrument Timbre Preference Test and the Intermediate Measures of Music Audiation

	Intermediate Measures of Music Audiation		
	Tonal	Rhythm	Composite
Timbre			
A	-.04	.00	-.02
B	.04	.10	.12
C	.10	.07	.10
D	-.05	-.02	-.05
E	-.06	-.06	-.06
F	.01	-.04	-.01
G	.01	-.09	-.05

aptitude are taken into account, students appear not to prefer or to dislike any timbre more or less than any other.

A finding related to music aptitude, though not directly related to the purpose or problems of the study, cannot be ignored. The levels of music aptitude of the fourth grade students who enrolled in beginning instrumental music instruction and remained in the program for two years are almost exactly the same as the levels of music aptitude of all fourth grade students in the school district. *Intermediate Measures of Music Audiation* means and standard deviations in Table 9 for the 57 students may be compared to corresponding means in Table 7 for the 258 students. Data available for many music programs across the country are no different. Thus it can be said with certainty that there is a great waste of human potential in terms of music aptitude in the schools. Approximately fifty percent of students who score in the upper twenty percent on valid music aptitude tests do not participate in special school music instruction. For whatever the reason or reasons, it would seem that little attempt is made by teachers and administrators to identify those students.

Table 9
Intermediate Measures of Music Audiation Means and Standard Deviations for Students Who Completed Two Years of Instrumental Music Instruction

	Mean	SD
Tonal	34.3	2.36
Rhythm	32.2	3.00
Composite	66.5	4.35

The five teachers who rated the instrumental performance achievement of the 57 students after they had received two years of instrumental instruction were given specific directions for determining the ratings. In each group, approximately 60 percent of students were to be given a rating of 3 (average), approximately 15 percent were to be given a rating of 2 (good) and 15 percent were to be given a rating of 4 (fair), and approximately 5 percent were to be given a rating of 1 (excellent) and approximately five percent were to be given a rating of 5 (poor) for tonal achievement, rhythm achievement, and overall musicianship. Each student received three ratings.

The means and standard deviations of the ratings are presented in Table 10. Although the means are slightly higher than desirable, the standard deviations represent sufficient variability for purposes of correlation. What is less than desirable, however, are the extremely high intercorrelations among the three ratings reported in Table 11. It seems that many teachers and performing musicians find it difficult to distinguish between students' tonal and rhythm achievement. The negative implications for instruction are enormous.

If students who learn to play an instrument for which they have a timbre preference as indicated by results on the *Instrument Timbre Preference Test* achieve more than students who learn to play an instrument for which they do not have a timbre preference as indicated by results on the *Instrument Timbre Preference Test*, that would support the predictive validity of the test. It can be seen in Table 12 that indeed the 10 students who studied instruments that are consistent with their timbre preferences achieve at a higher level than do the 47 students who studied instruments that are not consistent with their timbre preferences. The means for the former group are almost half a point or more (on a five point scale) above the averages (see Table 10), whereas the means for the latter group are one tenth of a point below the averages.

Because there were only ten students who elected to study instruments that were consistent with their timbre preferences, comparisons of the *Intermediate Measures of Music Audiation* in conjunction with the *Instrument Timbre Preference Test* to predict the performance achievement of students who study instruments that are consistent or inconsistent with their timbre preferences were not undertaken. However, correlations between students' music aptitude test scores and the teachers' instrumental performance achievement ratings, regardless of whether the students were playing instruments that were or were not consistent with their timbre preferences were computed. They are reported in Table 13. The magnitudes and relationships of the coefficients are similar to those found in previous longitudinal studies. Using the *Intermediate Measures of Music Audiation* as a basis, it can be said that more than 30 percent of the reason or reasons that students are successful in beginning instrumental music instruction is associated solely with their overall music aptitude.

Table 10
Means and Standard Deviations of the Teachers' Instrumental Music Performance Ratings

	Mean	SD
Tonal Achievement	3.2	1.21
Rhythm Achievement	3.4	1.11
Overall Musicianship	3.3	1.13

Table 11
Intercorrelations Among the Teachers' Instrumental Music Performance Ratings

	Tonal Achievement	Rhythm Achievement	Overall Musicianship
Tonal Achievement		.88	.96
Rhythm Achievement	.88		.93
Overall Musicianship	.96	.93	

There has been speculation about whether students' preferences on the *Instrument Timbre Preference Test* influence their instrumental music performance achievement as measured by teachers' ratings. With the possible exception of Timbre B, for which the results probably occurred by chance, the data in Table 14 suggest that students' tonal achievement, rhythm achievement, and overall musicianship follow the same patterns regardless of their specific timbre preference or preferences or of no preference at all. That finding, which seems logical, theoretically supports the construct validity of the test.

Table 12
Means and Standard Deviations of Teachers' Ratings of the Music Performance Achievement of Students Who Study Instruments that are Consistent or Inconsistent with Their Timbre Preferences

	Consistent Timbre Preference and Performance Medium		Inconsistent Timbre Preference and Performance Medium	
	Mean	SD	Mean	SD
Tonal Achievement	3.8	.63	3.1	1.19
Rhythm Achievement	3.8	.67	3.3	1.19
Overall Musicianship	3.7	.67	3.2	1.24

Table 13
Two-Year Longitudinal Predictive Validity Coefficients for the Intermediate Measures of Music Audiation

	Teachers' Ratings of Students' Instrumental Music Performance Achievement		
	Tonal Achievement	Rhythm Achievement	Overall Musicianship
Intermediate Measures of Music Audiation			
Tonal	.38	.43	.45
Rhythm	.40	.46	.44
Composite	.49	.54	.55

Table 14
Means and Standard Deviations of Students' Instrumental Music Performance Achievement According to Their Instrument Timbre Preferences

	Tonal Achievement		Rhythm Achievement		Overall Musicianship	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Timbre						
A (N=19)	3.3	1.09	3.6	1.00	3.5	1.00
B (N=3)	1.7	1.50	1.5	1.29	1.7	1.50
C (N=8)	3.0	1.41	3.3	1.18	3.1	1.24
D (N=17)	3.2	1.36	3.5	1.38	3.4	1.42
E (N=8)	3.3	1.41	3.5	1.23	3.4	1.42
F (N=4)	3.2	1.00	2.7	1.25	3.0	1.41
G (N=8)	3.1	1.12	3.0	1.30	2.7	1.38
No Preference (N=11)	2.9	1.00	3.0	1.04	3.0	1.00

It has been suggested that teachers' judgments of students' achievement in instrumental performance may be related to the instrument that the student is studying. That issue could be only indirectly addressed in this study by comparing how teachers evaluate the comparative success of students in beginning instrumental music who have different instrument timbre preferences. It can be seen in Table 15 that the relationship between the two factors is practically nil. The highest coefficient, -.14, indicates that there is, at most, less than two percent in common between teachers' achievement ratings and students' timbre preferences. The results are no different when the students' instrument timbre preferences are grouped into two categories of woodwind and brass. The mean differences, reported in Table 16, are negligible.

Table 15
Correlations Between Students' Instrument Timbre Preferences and Teachers' Ratings of Their Instrumental Performance Achievement

Timbre	Tonal Achievement	Rhythm Achievement	Overall Musicianship
A	.01	-.03	.00
B	-.08	-.02	-.07
C	.06	.12	.09
D	.06	-.01	.00
E	.10	-.02	.02
F	-.02	.00	.00
G	-.14	-.05	-.07

Table 16
Means and Standard Deviations of Teachers' Ratings of Students' Instrumental Performance Achievement According to Students' Woodwind or Brass Timbre Preferences

	Tonal Achievement		Rhythm Achievement		Overall Musicianship	
	Mean	SD	Mean	SD	Mean	SD
Woodwind (N=49) Timbres A, B, C, and D Combined	3.2	1.24	3.4	1.71	3.3	1.21
Brass (N=20) Timbres E, F, and G Combined	3.1	1.09	3.3	1.28	3.2	1.27

Conclusions

On the basis of the information derived from this study, it may be concluded that the theoretical validity of the *Instrument Timbre Preference Test* thus far supports its demonstrated practical validity, particularly in terms of longitudinal prediction. Further, it seems clear that some traditional beliefs and practices in education, specifically in music education, need to be reexamined. Perhaps continuing education for music teachers should become mandatory.

A limitation of the study, because of the small number of students who typically chose to enter and remain in beginning instrumental music programs, is that not every instrument that is represented by a synthesized timbre in the test was studied by students. Also, there were not enough students who received the same score for two or more timbres on the test to investigate how such a result might be best interpreted. Future research might be directed toward answering 1) whether a score of 9 or higher is sufficient for indicating a timbre preference for any one of the brass instruments associated with Timbres E, F, and G and 2) whether other instruments, such as soprano and tenor saxophones and all stringed instruments, might be associated with any of the existing synthesized timbres. The extent to which scores on the *Instrument Timbre Preference Test* change after instrumental music instruction, or simply as students become older, should be addressed in an extensive specific investigation. Information pertaining to such concerns, in conjunction with the results reported in this study and the information found in the test manual, could have a profound impact upon important aspects of music education.

Bibliography

Abeles, H. F. and S. Y. Porter. (1978). The Sex-Stereotyping of Musical Instruments. *Journal of Research in Music Education*, 26, 65-75.

Bernier, J. J. and R. E. Stafford. (1972). The Relationship of Musical Instrument Preference to Timbre Discrimination. *Journal of Research in Music Education*, 20, 283-285.

Coffman, D. D. and K. H. Sehmann. (1989). Musical Instrument Preference: Implications for Music Educators. *Update*, 7, 32-34.

Cutietta, R. A. and M. Foustalieraki. (1990). Preferences for Select Band and Non-Band Instrument Timbres Among Students in the United States and Greece. *Council for Research in Music Education*, 105, 72-80.

Geringer, J. M. (1977). Assessment of Children's Musical Instrument Preferences. *Journal of Music Therapy*, 14, 172-179.

Gordon, E. E.(1979). *Intermediate Measures of Music Audiation.* Chicago, GIA.

Gordon, E. E. (1984). *Instrument Timbre Preference Test.* Chicago, GIA.

Gordon, E. E. (1986). Final Results of a Two-Year Longitudinal Predictive Validity Study of the Instrument Timbre Preference Test and the Musical Aptitude Profile. *Council for Research in Music Education*, 89, 8-17.

Gordon, E. E. (1989). *A Two-Year Longitudinal Predictive Validity Study of the Instrument Timbre Preference Test and the Intermediate Measures of Music Audiation.* Predictive Validity Studies of IMMA and ITPT, GIA Monograph Series. Chicago, GIA.

Hufstader, R. A. (1974). Predicting Success in Instrumental Music Through the Use of Selected Tests. *Journal of Research in Music Education*, 22, 52-56.

Schmidt, C. P. and B. E. Lewis. (1988). A Validation Study of the Instrument Timbre Preference Test. *Psychology of Music*, 16, 143-155.

Weaver, S. C. (1987). An Investigation of the Relationships Between Preferences for Natural and Synthesized Timbres. *Dissertation Abstracts International*, 48, 864A. (University Microfilms No. 8716498).

Instrumental Performance Rating Scale

Name of Student _____

Instrument _____

School _____ Teacher _____

This is a five-point rating scale.

1 = Excellent
2 = Good
3 = Average
4 = Fair
5 = Poor

Comparing this student to all others in his or her class of beginning instrumental music students, please give him or her a rating of 1 to 5 in each of the categories listed below.

Tonal Achievement _____
Rhythm Achievement _____
Overall Musicianship _____

(Approximately 60% should be given a 3, 15% should be given a 4, 15% should be given a 2, 5% should be given a 1, and 5% should be given a 5.)

TAKING ANOTHER LOOK AT SCORING THE ADVANCED MEASURES OF MUSIC AUDIATION: THE GERMAN STUDY

Introduction

The words "talent," "ability," and gifted" are often used in a general way by music educators to describe musicality. The three words are as misleading as the word "musicality" itself. None of the words have specific meaning. All of them include, and thereby confuse, aptitude and achievement.

Music aptitude is a measure of a student's potential to achieve in music. Music achievement is a measure of what a student has actually achieved. Students with high music aptitude are able to learn from what they have learned to a much greater extent than are students with low music aptitude. The *concealed* (hitherto unknown) music aptitudes of many students, regardless of whether they have received instruction in music, are quickly *revealed* by a valid music aptitude test. A valid music aptitude "hears" what even a good music teacher does not "see." Of course students who score high on a valid music achievement test also score high on a valid music aptitude test. The reverse, however, is not necessarily true. Not all students who score low on a valid music achievement test also score low on a valid music aptitude test. Unfortunate as it may be, the truth is that students with low music achievement are often denied admission to a music degree program in a college or university without ever being tested for music aptitude. It is not unusual to find that some students with the highest levels of music aptitude demonstrate what is considered to be the lowest levels of music achievement. Given appropriate instruction, students with high music aptitude who initially demonstrate low music achievement generally will perform as well and possibly even better than students who have low music aptitude, the advantage of the earlier music instruction of the higher achievers being to little avail.

Objective music aptitude test scores should not be used in place of teachers' subjective judgments, nor should teachers' subjective judgments exclude the use of objective music aptitude test scores. To use one without the other is to know less about a student than teachers should know. Test scores can never replace the value of the opinions

of teachers, nor can the opinions of teachers ever replace the value of test scores. Both make unique contributions to the evaluation procedure. Thus the use of objective test scores and subjective opinions together create the most ideal circumstances for teaching and counseling a student. Perhaps the greatest value of a test score occurs when it is in disagreement with the opinion of a teacher. The resolution of such a disagreement should result in a better understanding of the student and improved instruction for the student.

It is not uncommon for an auditioning committee to deny students admission to a college or university music program simply because they may not sing or perform well on an instrument or because they may not have received a passing grade on a written test of music notation and music theory. Without also having those students' scores on a valid music aptitude test, such decisions may prove to be unwise. Some may have greater potential to profit from music instruction than others who were admitted. On the other hand, there are students who may sing or perform well on an instrument as a result of repetitive practice and imitation but, nevertheless, are unable to generalize what they have memorized to the learning of new compositions. Scores on a valid music aptitude test make an auditioning committee immediately aware of students who are uncertain in making musical inferences. Being able to generalize what is known to that which is unknown is the basis of true music achievement.

More important than information for making admission decisions is information that provides for adapting music instruction to students' individual musical differences after they are accepted into a music degree program. A valid music aptitude test will diagnostically indicate, for example, whether a student has stronger tonal aptitude than rhythm aptitude or stronger rhythm aptitude than tonal aptitude. Given that information before instruction begins, a teacher can prevent the loss of valuable time by immediately attending to the weaker of the two aptitudes without sacrificing achievement in the stronger of the two. Classroom and applied music teachers can individualize instruction through special assignments, thereby not frustrating students with low aptitude or not boring students with high aptitude. It must not be forgotten that all students are "musically gifted" to some extent. Each has at least some music aptitude and can profit from instruction in his own way. It is a teacher's responsibility

to know about a student's different music aptitudes and the comparative strengths and weaknesses of each.

The first standardized music aptitude test was published in 1919.[1] Neither that test nor any that has followed it was designed for use with college and university students. Though attempts have been made to use music aptitude tests that were originally written for students in grades four through twelve with younger and older students, none has proved successful.[2] In 1979, a music aptitude test became available for kindergarten children.[3] It was not until seventy years after Seashore's first test was published that a music aptitude test specifically designed for three and four year old children[4] and a music aptitude test specifically designed for undergraduate and graduate college and university music and non-music majors[5] were published. It is with the second of those two recent tests, the *Advanced Measures of Music Audiation*, that this paper is concerned.

Audiation

Before describing the *Advanced Measures of Music Audiation*, a brief explanation of audiation is in order.[6] Audiation is the basis of music aptitude. Thus it becomes the basis of music achievement. To audiate is to hear and comprehend music for which the sound is no longer or may never have been physically present. Audiation is to music what thinking is to language. Because the ability to imitate represents only the readiness to learn how to audiate, audiation is different from imitation. To imitate is simply to perform vocally or on an instrument without musically understanding what is being performed. For example, persons may imitate what was heard just a few moments ago and nonetheless be unaware of the tonality and the meter of what they are imitating, and they may not know that a modulation has taken place in the music. Audiation is also different from memorization. Persons may memorize music through imitation or from notation without audiating what they are performing. Unless

1. Carl E. Seashore, *Seashore Measures of Musical Talent* (New York: Columbia Phonograph Co., 1919).
2. For example, see Charles Harrington, "An Investigation of the Primary Level Musical Aptitude Profile for Use with Second and Third Grade Students," *Journal of Research in Music Education*, 17 (1969), pp. 193-201 and Robert Lee, "An Investigation of the Use of the Musical Aptitude Profile with College and University Freshman Music Students," *Journal of Research in Music Education*, 15 (1967), pp. 278-288.
3. Edwin E. Gordon, *Primary Measures of Music Audiation* (Chicago: GIA, 1979).
4. Edwin E. Gordon, *Audie* (Chicago: GIA, 1989).
5. Edwin E. Gordon, *Advanced Measures of Music Audiation* (Chicago: GIA, 1989).
6. A detailed explanation of the types and hierarchical stages of audiation may be found in Chapter Two of *Learning Sequences in Music: Skill, Content, and Patterns* by Edwin E. Gordon. (Chicago: GIA, 1988).

one can audiate what is being imitated or has been memorized, he most likely will perform at best with poor intonation, inaccurate rhythm, and limited expression. Moreover, the recognition of music has as little to do with audiation as does the imitation and memorization of music.

What is imitated or memorized without audiation is quickly forgotten. What is improvised, created, or recalled as a result of audiation may be remembered for a long time. Further, unless an instrumentalist or vocalist can notationally audiate (hear what is seen in notation before it is performed), he will be unable to bring meaning to notation as he attempts to read it; he will be able only to recite the letter names and time value names of the notes and to theoretically define signs and symbols seen in notation. The word "denote" applies simply to the reading of notation. The word "connote" applies to the audiation of notation. Unless a performer can audiate, he will not have the ability to improvise or create.

Description of the Advanced Measures of Music Audiation

The *Advanced Measures of Music Audiation* is a cassette recorded music aptitude test for college and university students that requires approximately fifteen minutes of administration time. It may be administered to individuals or to groups of students. In addition to the thirty test questions, each containing a pair of short musical phrases, directions for taking the test and practice exercises are also included on the cassette. Each test question is programmed on an Apple Macintosh computer and performed by a professional musician on a Yamaha DX-7 synthesizer. The original music was composed specifically for the test.

Formal music achievement is not a requirement for taking the *Advanced Measures of Music Audiation.* Regardless of whether a student can play an instrument, sing, or read notation, or has taken courses in music theory, he may score high on the test. On the other hand, because a student is a performer, can read notation, or is familiar with music theory is no guarantee that he will score high on the test. The student simply indicates, by filling a space on the answer sheet, 1) whether the two phrases that are heard sound the same, 2) whether they sound different because of a tonal change, or 3) whether they sound different because of a rhythm change. The space is filled

in the "tonal" column if there is a tonal change in the second part, the space is filled in the "rhythm" column if there is a rhythm change in the second part, or the space is filled in the "same" column if the second part is the same as the first part. There is only one correct answer for each question. Because of the unique process for scoring the test, which will be described in detail later, the *Advanced Measures of Music Audiation* yields three scores: a *Tonal* score; a *Rhythm* score; and a *Total* score, which is a composite of the two.

The amount of time the student is given to decide upon the answer to each question is highly related to audiation. If too much time is allowed, memorization through imitation intervenes. If too little time is allowed, there is not enough silent time to audiate. Just enough time must be given to the student to summarize and generalize the essential aspects of the pair of patterns so that audiation may take place. One summarizes and generalizes in audiation while listening to sounds in music in a manner similar to the way one summarizes and generalizes in thinking while listening to words in conversation.

A representative sample of more than 5,000 college and university students from 54 institutions in 27 states participated in the national standardization of the *Advanced Measures of Music Audiation*. Of the total sample, 3,206 were undergraduate and graduate music majors and 2,130 were undergraduate and graduate non-music majors.[7] Specific norms are presented in the test manual.

Background of Research

A great deal of research was undertaken during the development of the *Advanced Measures of Music Audiation*. In addition to means, standard deviations, item discrimination and difficulty indexes, split-halves and retest reliabilities, standard errors of measurements and differences, and intercorrelations being established, preliminary validity for the test was investigated.[8] One year after the publication of the test, a study of its predictive validity was completed and published.[9] It is described below. More important to the research

---

7. A full description of the standardization program, which followed sampling procedures based upon demographic information contained in the 1986-87 Heads Music Data Summaries of the Higher Education Arts Data Service (HEADS), may be found in Part 7, pages 36 through 40, of the Manual for the *Advanced Measures of Music Audiation*.

8. For a full account, see pages 40 through 51 of the Manual for the *Advanced Measures of Music Audiation*.

9. Edwin E. Gordon, *Predictive Validity Study of AMMA: A One-Year Longitudinal Predictive Validity Study of the Advanced Measures of Music Audiation*, GIA Monograph Series (Chicago: GIA, 1989).

presented in this paper, however, is an explanation of the unique scoring procedure used for the *Advanced Measures of Music Audiation*. The predictive validity study will be described and then an explanation of the scoring procedure will follow.

Predictive Validity of the Advanced Measures of Music Audiation

At the beginning of the 1989 fall semester, the *Advanced Measures of Music Audiation* was administered to all 225 undergraduate and graduate members of the orchestra, concert choir, and band of the Boyer College of Music of Temple University in Philadelphia, Pennsylvania. There were 90 students in the orchestra, 46 in the concert choir, and 89 in the band. Each group included freshman through doctoral students.

At the end of the the 1990 spring semester, all students to whom the *Advanced Measures of Music Audiation* had been administered in the fall of 1989 and who were taking jury examinations were asked to perform a short etude that they had time to prepare. The etude constituted the validity criterion in the investigation.[10] To equalize the difficulty of the etude for all students to the greatest extent possible, it was specially composed and written in an appropriate key for every orchestral and band instrument, every voice, and piano.

Of the 225 students who took the *Advanced Measures of Music Audiation* in 1989, slightly more than half, 114, recorded the etude in 1990. The reasons for the decrease in numbers are 1) approximately one third of the students were not music majors, and thus they did not take jury examinations during the time the etudes were performed, 2) other students performed recitals and therefore were not required to take jury examinations, 3) percussion students did not take jury examinations at the regularly scheduled times, and 4) some students who did take jury examinations exercised their rights under the HRS Regulations (Protection of Human Research Subjects) and chose not to record the etude. Nevertheless, the scores of the 114 students provided adequate variability for the statistical analyses to be undertaken.

Using a three dimensional rating scale of five points each, three judges independently evaluated the students' recorded performances

10. It was found in the standardization program of the *Advanced Measures of Music Audiation* that neither chronological age nor years of music training are related to scores on the test. With regard to the etude that served as the validity criterion, it is technically simple enough, yet nonetheless revealing of one's tonal and rhythm capabilities, for even the youngest student to learn to perform.

of the etude. The same rating scale was used with voice students and students who played a band or orchestral instrument. A separate rating scale was used with piano students.

The most important correlation was that between the *Total* score on the *Advanced Measures of Music Audiation* and the total score of the three judges' combined ratings of the etude performances. It was found to be quite high: .82. In simplified terms, that coefficient indicates that more than 67 percent of the reason or reasons for college students' success in music can be predicted by *Total* test scores on the *Advanced Measures of Music Audiation.*

Scoring Procedure for the Advanced Measures of Music Audiation

The logic of the scoring procedure for the *Advanced Measures of Music Audiation* is based on the belief that if, in a test question, the two parts are the same or if there is a rhythm change in the second part, and the student makes an error by indicating that there was a tonal change in the second part, one point should be subtracted from his *Tonal* score. For a student to think that a tonal change took place when in fact it did not is as relevant to his tonal aptitude as thinking that a tonal change did not take place when in fact it did. Similarly, when a student erroneously indicates that a rhythm change took place in the second part, that is relevant to his rhythm aptitude and one point should be subtracted from the student's *Rhythm* score.

When a student does not mark an answer for a test question of which he is unsure, no credit is given for a correct answer and nothing is subtracted from his score. It is reasoned that if a student knows that he does not know the correct answer to a question, he knows more than a student who does not know that he does not know the correct answer to a question, and therefore gives an incorrect answer. Nothing is subtracted as a result of answers being omitted or being marked in the "same" column.

The *Advanced Measures of Music Audiation* includes thirty questions. Ten questions incorporate a tonal change in the second part, ten questions incorporate a rhythm change in the second part, and both parts are the same for the remaining ten questions. The *Tonal* score consists of the total of correct answers for the ten questions with a tonal change plus the ten questions in which both parts are the same. That is the student's raw (unadjusted) *Tonal*

score. Then the number of incorrect tonal answers are subtracted from that score. To avoid negative scores, a constant is added to a student's raw (unadjusted) score. By adding a constant of twenty points, a student may receive a maximum of forty points on the *Tonal* test. That represents the student's adjusted *Tonal* score. The *Rhythm* score consists of the total of correct answers for the ten questions with a rhythm change plus the ten questions in which both parts are the same. That is the student's raw (unadjusted) *Rhythm* score. Then the number of incorrect rhythm answers are subtracted from that score. By adding a constant of twenty points, a student may receive a maximum of forty points on the *Rhythm* test. That represents the student's adjusted *Rhythm* score. A perfect adjusted score on the *Total* test is eighty points.

The literature is replete with admonishments against correcting test scores for guessing. There is no evidence that the validity of a student's test score will increase if that student is penalized by the loss of points for guessing or rewarded by the addition of points for declining to guess. It is known that the correlation between students actual test scores and their corrected-for-guessing test scores is extremely high, and in most cases perfect.

The adjusted scores on the *Advanced Measures of Music Audiation* should not be construed as being corrected-for-guessing scores. The scores are adjusted on the basis of how well a student directly and indirectly audiates tonally and rhythmically. To simply and without qualification correct a score for guessing by adding or subtracting a point for a given answer is quite different from and less valid than specifically to subtract a point because a student believed that he heard a rhythm change or a tonal that did not exist.

It was discovered in the standardization program of the *Advanced Measures of Music Audiation* that the unadjusted and adjusted *Tonal* scores correlate .89, the unadjusted and adjusted *Rhythm* scores correlate .87, and the unadjusted and adjusted *Total* test scores correlate .93. The reliabilities of the unadjusted *Tonal*, *Rhythm*, and *Total* scores are .76, .76, and .80, respectively. In comparison, the reliabilities of the adjusted *Tonal*, *Rhythm*, and *Total* scores are .89, .90, and .92, respectively. As would be expected, negative correlations were found between the number of points to be subtracted from the raw (unadjusted) scores and the adjusted scores.

Purpose and Problems of the Study

The purpose of the current study is to review and to gather more specific information about the scoring procedure for the *Advanced Measures of Music Audiation*. The research problems were 1) to determine whether the heretofore statistical analyses favoring the adjusted score over the unadjusted score remain consistent for students in a European country, 2) to investigate more detailed aspects of the comparative means, standard deviations, and reliabilities of the unadjusted and adjusted scores as well as of the adjustment score (the number of points subtracted) itself, and 3) to establish the correlations among the unadjusted, adjusted, and adjustment scores.

Design of the Study

During the 1990 fall semester, 129 students who were attending the Hochschule in Bremen, Germany were administered the *Advanced Measures of Music Audiation*, in English, in accordance with the standard procedures outlined in the test manual.[11] The age range of the students was comparable to that of undergraduate through graduate students in the United States. Of the 129 students, 57 were majoring in music and 72 were majoring in education. The majority of students who were not majoring in music were intending to become art or physical education teachers.

For the first analysis, three *Tonal* scores, three *Rhythm* scores, and one *Total* score were derived for each student. The data are presented in Table 1 for the music majors and in Table 2 for the non-music majors.

The "Tonal 1" score is the tonal raw (unadjusted) score, the number of correct answers for the ten questions in which there was a tonal change in the second part plus the number of correct answers for the ten questions in which the two parts are the same. A maximum "Tonal 1" score is 20. The "Rhythm 1" score is the rhythm raw (unadjusted) score, the number of correct answers for the ten questions in which there was a rhythm change in the second part plus the number of correct answers for the ten questions in which the two parts are the same. A maximum "Rhythm 1" score is 20.

The "Tonal 2" score is the tonal adjustment score, the number of points subtracted from the tonal raw (unadjusted) score for incorrect answers. If a student indicated that a tonal change took place in the

11. The writer is grateful to Prof. Dr. Gunter Kleinen of the Hochschule for overseeing the administration of the test.

second part of a question but in fact it did not, one point was subtracted from the tonal raw (unadjusted) score. Considering the ten questions in which a rhythm change takes place in the second part and the ten questions in which both parts are the same, a maximum "Tonal 2" score is 20. The "Rhythm 2" score is the rhythm adjustment score, the number of points subtracted from the rhythm raw (unadjusted) score for incorrect answers. If a student indicated that a rhythm change took place in the second part of a question but in fact it did not, one point was subtracted from the rhythm raw (unadjusted) score. Considering the ten questions in which a tonal change takes place in the second part and the ten questions in which both parts are the same, a maximum "Rhythm 2" score is 20.

Table 1
Means, Standard Deviations, and Score Ranges for German Music Majors on the Advanced Measures of Music Audiation

	Mean	Standard Deviation	Score Range
Tonal 1 (Raw Score)	13.1	2.43	7-18
Tonal 2 (Adjustment Score)	5.0	1.63	1-8
Rhythm 1 (Raw Score)	12.5	2.28	8-18
Rhythm 2 (Adjustment Score)	1.4	1.37	0-7
Tonal (Adjusted Score)	28.4 (28.3)*	3.37 (4.12)	20-36
Rhythm (Adjusted Score)	31.2 (30.8)	3.04 (3.52)	21-38
Total (Adjusted Score)	59.6 (59.1)	6.40 (7.41)	45-73

* Means and standard deviations in parenthesis were derived from the American standardization program.

Table 2
Means, Standard Deviations, and Score Ranges for German Non-Music Majors on the Advanced Measures of Music Audiation

	Mean	Standard Deviation	Score Range
Tonal 1 (Raw Score)	10.8	3.42	2-17
Tonal 2 (Adjustment Score)	5.7	2.18	1-11
Rhythm 1 (Raw Score)	11.0	2.86	5-19
Rhythm 2 (Adjustment Score)	2.7	1.87	0-10
Tonal (Adjusted Score)	25.6 (24.3)*	4.57 (4.89)	14-35
Rhythm (Adjusted Score)	28.8 (27.4)	4.02 (4.11)	17-38
Total (Adjusted Score)	54.2 (51.7)	8.09 (8.49)	34-73

* Means and standard deviations in parenthesis were derived from the American standardization program.

The "Tonal" score is the adjusted score, the actual score that a student receives on the *Tonal* test after a constant of 20 points is added to the tonal raw score and the points for the adjustment score are subtracted from that sum. A maximum *Tonal* score is 40. The "Rhythm" score is the adjusted score, the actual score that a student receives on the *Rhythm* test after a constant of 20 points is added to the rhythm raw score and the points for the adjustment score are subtracted from that sum. A maximum *Rhythm* score is 40. The "Total" score is the adjusted score, the actual score, which is the sum of the *Tonal* and *Rhythm* test scores, that a student receives on the *Total* test.

Included in Tables 1 and 2, along with the means and standard deviations for the raw (unadjusted) scores, adjustment scores, and adjusted scores, are the score ranges for those scores. For purposes of comparison, the means and standard deviations for the actual (adjusted) scores on the *Tonal* test, the *Rhythm* test, and the *Total* test are given in parenthesis.

It can be seen in Table 1 that the means and standard deviations for the *Tonal*, *Rhythm*, and *Total* tests are highly similar for German and American music students. Though of little practical signifi-

cance, the means for the German non-music students on the *Tonal*, *Rhythm*, and *Total* tests reported in Table 2 are higher, less than only a point and a half for the *Tonal* and *Rhythm* tests, than those for the American students. Because the differences are so small, they are not necessarily reflected in the corresponding standard deviations.

A comparison of the data in Tables 1 and 2 indicate that, as with American students, German students who are music majors overall score higher on the *Advanced Measures of Music Audiation* than do German students who are non-music majors. The differences between American music and non-music majors is greater, however, than that between German music and non-music majors. Of particular interest is that the highest score for German music majors is no higher than the highest score for German non-music majors. The situation is exactly the same for American students on the *Rhythm* test. With regard to the *Tonal* test, the difference, favoring the music majors, is only one point. That the score ranges are similar for music and non-music majors but the means are higher for music majors than non-music majors in both countries indirectly supports the assumption that the *Advanced Measures of Music Audiation* is more a music aptitude test than a music achievement test. It would seem that the reason music majors score higher than non-music majors has little to do with music achievement. If it did, the mean differences between the two groups would be much greater. The reason the means are slightly higher for music majors is probably that more persons with high music aptitude are naturally compelled to study music than are persons with low music aptitude.

The raw (unadjusted) score means, as seen in Tables 1 and 2, are higher for the music majors than for the non-music majors. The tonal adjustment score means are nearly the same. The rhythm adjustment score mean for the non-music majors, however, is almost twice as high that for the music majors. It would seem that non-music majors are less certain about what they interpret as constituting a rhythm change than what they interpret as constituting a tonal change.

Analyses of the reliability characteristics of the *Advanced Measures of Music Audiation* were undertaken separately for music majors and for non-music majors and for both groups combined. Because of the unique scoring procedure for the test, which results in adjusted scores, the derivation of split-halves reliabilities also requires a unique procedure. The following is a detailed description of that

procedure: 1) The 20 questions on the *Tonal* test (tonal raw unadjusted score) and the 20 questions on the *Rhythm* test (rhythm raw unadjusted score) were scored. 2) The item difficulty levels and discrimination values were computed separately for each test. 3) Each test was divided into "equal" halves, 10 questions in each, by matching questions to the extent possible in terms of their difficulty levels and discrimination values. 4) The two halves of the *Tonal* test were scored and the two halves of the *Rhythm* test were scored. 5) A constant of 10 was added to each half of the *Tonal* test and a constant of 10 was added to each half of the *Rhythm* test. 6) The two halves of the Tonal test were scored again to derive two tonal adjustment scores and the two halves of the *Rhythm* test were scored again to derive two rhythm adjustment scores. 7) For each student, the tonal adjustment score for one half of the test was subtracted from the *Tonal* test score for one half of the test, and the tonal adjustment score for the other half of the test was subtracted from the *Tonal* test score for the other half of the test. The resulting two scores for each set were correlated and corrected for length through the use of the Spearman-Brown Prophecy Formula. That coefficient represents the split-halves reliability of the adjusted score for the *Tonal* test. 8) For each student, the rhythm adjustment score for one half of the test was subtracted from the *Rhythm* test score for one half of the test, and the rhythm adjustment score for the other half of the test was subtracted from the *Rhythm* test score for the other half of the test. The resulting two scores for each set were correlated and corrected for length through the use of the Spearman-Brown Prophecy Formula. That coefficient represents the split-halves reliability of the adjusted score for the *Rhythm* test. 9) The split-halves reliability of the *Total* test was derived by summing the adjusted scores for one half of the *Tonal* test and one half of the *Rhythm* test and by summing the adjusted scores of the other half of the *Tonal* test and the other half of the *Rhythm* test. The two sums were correlated and corrected for length through the use of the Spearman-Brown Prophecy Formula. The means, standard deviations, and corrected split-halves reliability coefficients for the raw (unadjusted) scores and the adjusted scores on the *Tonal*, *Rhythm*, and *Total* tests are presented in Table 3 for the music majors, Table 4 for the non-music majors, and Table 5 for the music majors and non-music majors combined.

A comparison of Tables 3, 4, and 5 indicates that the central

tendency and variability for the corresponding first and second halves of each test within every table are quite similar. Considering the brevity of the test, however, the magnitude of the reliabilities of the raw (unadjusted) scores are not unrealistic. Nevertheless, the reliability coefficients of the adjusted scores are substantial, in all cases, particularly for the *Rhythm* test, higher than the raw (unadjusted) scores. Such a finding, which corroborates the results derived from the data based upon all students who participated in the national standardization program, suggests that the unique scoring procedure for all three tests which constitute the *Advanced Measures of Music Audiation* is appropriate.

It is interesting to note in Tables 3, 4, and 5 that the reliabilities of the adjusted scores are not materially different for the music majors, non-music majors, and both groups combined. Especially with regard to the *Total* test, combining the results of the music and non-music majors, and thereby artificially increasing the variability of the score distributions, has no appreciable effect on the reliability of either the unadjusted or adjusted scores. That is encouraging, because norms are provided in the *Advanced Measures of Music Audiation* test manual separately for music majors and non-music majors. Probably because of the homogeneity and the relatively small number of music majors, all reliabilities are somewhat lower for the music majors than for the non-music majors.

For theoretical interest, correlations were computed among the raw (unadjusted) scores, adjustment scores, and adjusted scores separately for the music majors and non-music majors. They are presented for the music majors in Table 6 and for the non-music majors in Table 7. The corresponding coefficients in the two tables for the two groups of students are almost identical. In terms of the psychological constructs underlying the *Advanced Measures of Music Audiation*, the data support the belief that the test emphasizes and is measuring music aptitude rather than music achievement.

As would be expected, the correlation between raw (unadjusted) scores and adjustment scores is negative. That is, the more questions a student answers correctly the fewer he answers incorrectly, or the more questions a student answers incorrectly the fewer he answers correctly. That, however, is more nearly the case for the *Tonal* test than for the *Rhythm* test. Though statistically significant, the correlations for the *Rhythm* test are not high enough to be of practical

Table 3
Means, Standard Deviations, and Split Halves Reliability Coefficients of Raw Scores and Adjusted Scores for German Music Majors on the Advanced Measures of Music Audiation

	Mean	Standard Deviation	Reliability*
Raw Scores			
Tonal - First Half	6.2	1.41	
Tonal - Second Half	6.9	1.34	.70
Rhythm - First Half	6.2	1.35	
Rhythm - Second Half	6.3	1.38	.52
Composite - First Half	12.4	2.50	
Composite - Second Half	13.2	2.60	.70
Adjusted Scores			
Tonal - First Half	4.1	1.87	
Tonal - Second Half	4.3	2.04	.83
Rhythm - First Half	5.6	1.63	
Rhythm - Second Half	5.6	1.61	.81
Composite - First Half	9.7	3.15	
Composite - Second Half	9.9	3.35	.91

* The coefficients are Spearman-Brown corrected.

Table 4
Means, Standard Deviations, and Split Halves Reliability Coefficients of Raw Scores and Adjusted Scores for German Non-Music Majors on the Advanced Measures of Music Audiation

	Mean	Standard Deviation	Reliability*
Raw Scores			
Tonal - First Half	5.2	1.96	
Tonal - Second Half	5.6	1.99	.71
Rhythm - First Half	5.7	1.53	
Rhythm - Second Half	5.3	1.67	.63
Composite - First Half	10.9	3.20	
Composite - Second Half	10.9	3.32	.77
Adjusted Scores			
Tonal - First Half	2.7	2.21	
Tonal - Second Half	2.9	2.42	.90
Rhythm - First Half	4.4	1.98	
Rhythm - Second Half	4.2	1.97	.88
Composite - First Half	7.1	3.98	
Composite - Second Half	7.1	4.16	.91

* The coefficients are Spearman-Brown corrected.

Table 5
Means, Standard Deviations, and Split Halves Reliability Coefficients of Raw Scores and Adjusted Scores for German Music and Non-Music Majors on the Advanced Measures of Music Audiation

	Mean	Standard Deviation	Reliability*
Raw Scores			
Tonal - First Half	5.7	1.81	
Tonal - Second Half	6.2	1.84	.74
Rhythm - First Half	5.9	1.46	
Rhythm - Second Half	5.8	1.63	.88
Composite - First Half	11.6	3.00	
Composite - Second Half	12.0	3.25	.78
Adjusted Scores			
Tonal - First Half	3.3	2.16	
Tonal - Second Half	3.5	2.35	.90
Rhythm - First Half	4.9	1.92	
Rhythm - Second Half	4.8	1.94	.90
Composite - First Half	8.2	3.84	
Composite - Second Half	8.3	4.07	.92

* The coefficients are Spearman-Brown corrected.

Table 6
Correlations Among Raw Scores, Adjustment Scores, and Adjusted Scores of German Music Majors on the Advanced Measures of Music Audiation

Tonal 1 and Tonal 2	-.68
Rhythm 1 and Rhythm 2	-.27
Tonal 1 and Tonal	.94
Rhythm 1 and Rhythm	.85
Tonal 1 and Rhythm 1	.79
Tonal 2 and Rhythm 2	.16
Tonal and Rhythm	.72
Tonal and Total	.92
Rhythm and Total	.89
Tonal 2 and Tonal	-.88
Tonal 2 and Rhythm	-.59
Tonal 2 and Total	-.81
Rhythm 2 and Tonal	-.34
Rhythm 2 and Rhythm	-43
Rhythm 2 and Total	-.40

Tonal 1 = Tonal Raw Score
Rhythm 1 = Rhythm Raw Score

Tonal 2 = Tonal Adjustment Score
Rhythm 2 = Rhythm Adjustement Score

Tonal = Tonal Adjusted Score
Rhythm = Rhythm Adjusted Score
Total = Adjusted Composite Score (T and R combined)

importance. Why there is a discrepancy between the results for the *Tonal test* and the *Rhythm* is not readily apparent.

The high positive correlations between the raw (unadjusted) scores and the adjusted scores was also anticipated. It is interesting to remember that as high as the coefficients are, the adjusted scores encompass enough more precision to possess greater reliability than the less precise raw (unadjusted) scores. The magnitude of unique variance, though not exceptionally large, between the raw (unadjusted) score and adjusted score distributions would seem to suggest that the additional steps of adding a constant to the raw (unadjusted) scores and subtracting the adjustment scores to obtain the adjusted scores is justified.

The correlation between the raw (unadjusted) scores for the *Tonal* test and the *Rhythm* test is .79 for the music majors and .76 for the non-music majors. Those high correlations, which parallel those found in the analysis of the standardization data, are not surprising considering that the scores for both tests include the same ten questions that have 'Same" as the correct answer. That is, the tests share half the number of questions in each test. What is surprising is that the correlations are not higher. Nevertheless, even taking the test reliabilities into account, there appears to be enough unique variance in the score distributions associated with each test to provide for diagnostic interpretations. Certainly, a student can receive a high score on one test and a low score on the other, and vice versa. Of greater theoretical importance is that the adjusted *Tonal* and *Rhythm* test scores do not correlate any higher than do the *Tonal* and *Rhythm* raw (unadjusted) scores. In fact, as can be seen in Tables 6 and 7, the observed coefficients for the adjusted *Tonal* and *Rhythm* test scores are lower. They are .72 for the music majors and .71 for the non-music majors.

The only coefficients that are not statistically significant in Tables 6 and 7 are for the *Tonal* and *Rhythm* adjustment scores for the music majors and non-music majors. They indicate no relationship whatsoever between the two sets of scores. Students who incorrectly believe that a tonal change has occurred in the second part of some test questions do not systematically incorrectly believe that a rhythm change has occurred in the second part of other test questions. That finding helps explain why the correlation between the *Tonal* and *Rhythm* adjusted test scores is somewhat lower than the correlation between the *Tonal* and *Rhythm* raw (unadjusted) scores.

Table 7
Correlations Among Raw Scores, Adjustment Scores, and Adjusted Scores of German Non-Music Majors on the Advanced Measures of Music Audiation

Tonal 1 and Tonal 2	-.61
Rhythm 1 and Rhythm 2	-.28
Tonal 1 and Tonal	.91
Rhythm 1 and Rhythm	.84
Tonal 1 and Rhythm 1	.76
Tonal 2 and Rhythm 2	.08
Tonal and Rhythm	.71
Tonal and Total	.94
Rhythm and Total	.90
Tonal 2 and Tonal	-.82
Tonal 2 and Rhythm	-.56
Tonal 2 and Total	-.77
Rhythm 2 and Tonal	-.34
Rhythm 2 and Rhythm	-.57
Rhythm 2 and Total	-.47

Tonal 1 = Tonal Raw Score
Rhythm 1 = Rhythm Raw Score
Tonal 2 = Tonal Adjustment Score
Rhythm 2 = Rhythm Adjustement Score

A comparison of the correlations between the adjusted *Tonal* test score and the adjusted *Total* test score on the one hand and the adjusted *Rhythm* test score and the adjusted *Total* test score on the other indicates that the two subtests have a similar amount in common with the *Total* test score. Neither subtest, however, accounts for enough of the *Total* test variance to render the other unnecessary. Both contribute substantially to measuring a student's overall music aptitude.

Just as there are negative relationships between the adjustment scores and the raw (unadjusted) scores, so too are there negative relationships between the adjustment scores and the adjusted scores. That the negative correlations are higher between the adjustment scores and the adjusted scores than between the adjustment scores and the raw

(unadjusted) scores supports the need for the unique scoring procedure. In general and for whatever the reason, the negative coefficients in Tables 6 and 7 are higher between the *Tonal* test adjustment scores and the *Tonal* adjusted scores than between the *Rhythm* test adjustment scores and the *Rhythm* test adjusted scores.

The samples of German music majors and non-music majors to whom the *Advanced Measures of Music Audiation* was administered were comparatively small. For that reason, the percentile rank equivalents of the adjusted scores are not as broad as those for the American standardization populations. Specifically, the adjusted score distributions for the German students, though normally distributed, do not extend to the maximum obtainable on the tests. Neither do the adjusted scores for the German students extend as low as they do for the students who participated in the American standardization program. Nevertheless, for purposes of theoretical comparison as well as limited practical use, percentile rank equivalents for the adjusted *Tonal*, *Rhythm*, and *Total* scores for German music majors and non-music majors are reported in Table 8.

Conclusions

Considering that 1) the theoretical level of the validity of a test cannot be higher than the square root of its reliability and 2) the unique procedure for scoring the *Advanced Measures of Music Audiation* increases the reliability of the test as compared to the standard scoring procedure, it may be concluded that the scoring procedure for the *Advanced Measures of Music Audiation* should remain unchanged. The cross validation of American and German data and the additional ancillary information gathered in Germany support the design of the test as well as the use of the test in other countries that share a music culture similar to that of the United States.

Table 8
Percentile Ranks for German Music and Non-Music Majors on the Advanced Measures of Music Audiation

Tonal				Rhythm				Composite			
Major		Non-Major		Major		Non-Major		Major		Non-Major	
AS	PR	AS	PR	AS	PR	AS	PR	AS	PR	AS	PR
36	99	35	99	38	99	38	99	73	99	73	99
35	96	34	97	37	97	37	95	72	97	72	98
34	93	33	95	36	94	36	90	71	95	71	97
33	87	32	92	35	91	35	85	70	93	70	96
32	84	31	88	34	87	34	80	69	91	69	95
31	80	30	83	33	76	33	75	68	89	68	94
30	72	29	78	32	68	32	69	67	87	67	93
29	61	28	72	31	54	31	63	66	85	66	92
28	55	27	65	30	43	30	57	65	83	65	90
27	45	26	57	29	28	29	50	64	80	64	87
26	33	25	50	28	17	28	43	63	76	63	84
25	26	24	44	27	12	27	37	62	71	62	81
24	15	23	39	26	9	26	31	61	66	61	78
23	7	22	35	25	6	25	25	60	60	60	75
22	4	21	30	24	3	24	21	59	54	59	70
21	2	20	25	23	1	23	19	58	48	58	66
20	1	19	20			22	16	57	42	57	62
		18	15			21	13	56	37	56	58
		17	11			20	10	55	33	55	54
		16	7			19	7	54	29	54	50
		15	4			18	3	53	26	53	46
		14	1			17	1	52	23	52	43
								51	20	51	40
								50	17	50	37
								49	13	49	33
								48	10	48	29
								47	7	47	25
								46	4	46	22
								45	1	45	20
										44	18
										43	16
										42	14
										41	12
										40	10
										39	8
										38	6
										37	4
										36	3
										35	2
										34	1

AS = Adjusted Scores
PR = Percentile Ranks

Bibliography

Gordon, E. E. (1989). *Advanced Measures of Music Audiation.* Chicago: GIA.

Gordon, E. E. (1989). *Audie.* Chicago: GIA.

Gordon, E. E. (1989). *Predictive Validity Study of AMMA: A One-Year Longitudinal Predictive Validity Study of the Advanced Measures of Music Audiation.* Chicago: GIA.

Gordon, E. E. (1988). *Learning Sequences in Music: Skill, Content, and Patterns.* Chicago: GIA.

Gordon, E. E. (1986). *The Nature, Description, Measurement, and Evaluation of Music Aptitudes.* Chicago: GIA.

Gordon, E. E. (1979). *Primary Measures of Music Audiation.* Chicago: GIA.

Gould, S. J. (1981). *The Mismeasure of Man.* New York: W. W. Norton and Company.

Harrington, C. (1969). "An Investigation of the Primary Level Musical Aptitude Profile for Use with Second and Third Grade Students." *Journal of Research in Music Education,* 17, 193-201.

Lee, R. (1967). "An Investigation of the Use of the Musical Aptitude Profile with College and University Freshman Music Students," *Journal of Research in Music Education,* 15, 278-288.

Mursell, J. L. (1947). *Psychological Testing.* New York: Longmans, Green and Company.

Radocy, R. E. and J. D. Boyle. (1979). *Psychological Foundations of Musical Behavior.* Springfield, Illinois: Charles C. Thomas.

Seashore, C. E. (1919). *Seashore Measures of Musical Talent.* New York: Columbia Phonograph Company.

Shuter-Dyson, R. and C. Gabriel. (1981). *The Psychology of Musical Ability.* London: Methuen.

Thorndike, R. and E. Hagen. (1977). *Measurement and Evaluation in Psychology and Education.* New York: John Wiley and Sons.

G-3706 $11.95

7 85147 37060