Part 7

TECHNICAL ASPECTS OF THE ADVANCED MEASURES OF MUSIC AUDIATION

Procedures Used in Establishing the Norms

Overview of the Standardization Program

The *Advanced Measures of Music Audiation* was nationally standardized in the 1988-89 school year in a carefully planned and executed program. The standardization program is described in the following outline.

1. A representative sample of college and university music majors and non-music majors was selected through the use of sampling procedures based upon demographic information contained in the 1986-87 Heads Music Data Summaries of the Higher Education Arts Data Service (HEADS). The report was prepared for the National Association of Schools of Music (NASM).
2. The test was administered in the manner prescribed in this manual, using standard answer sheets.
3. The answer sheets were scored by the GIA Scoring Service.
4. Raw score frequency distributions were prepared.
5. Percentile norms were derived from distributions of raw scores for undergraduate and graduate music majors and non-music majors.
6. Data were accumulated on the relationship of music aptitude with 1) the musical status of students, 2) the graduate and undergraduate status of students, 3) the sex of students, 4) the geographical area of schools, 5) the public and private standing of schools, and 6) the size of schools.

The high school norms are based on a smaller, though comprehensive, sample of both music and non-music students. Approximately 15% of the students are musicians, the majority of those being members of school music performance ensembles.

Selection of the Standardization Sample

In establishing the norms of a test, the number of students tested is not so

important as the way in which the students were selected. To establish representative norms, it is crucial that the factors which might influence test performance be controlled. One of the most authoritative sources of demographic information pertaining to music in higher education is the Higher Education Arts Data Service. In order to undertake the standardization of the *Advanced Measures of Music Audiation,* the data contained in the 1986-87 HEADS report for NASM member institutions which grant music degrees were analyzed, simplified, and categorized by the test author. The results yielded the following information.

1. There were 409 member schools in 47 states, including Washington, D. C.
2. There were 234 (57.2%) public schools and 175 (42.8%) private schools.
3. Undergraduate music major enrollments for public and private institutions were different as described below.

Public Institutions		Private Institutions	
1 to 100	39.4%	1 to 50	49.54%
101 to 200	32.6%	51 to 100	22.2%
201 to 400	19.1%	101 to 200	15.2%
401 or more	8.9%	201 or more	13.1%

4. The institutions were located in six geographic areas in the following manner:
   - Northwest — 52 institutions (12.8%) including the 10 states of Connecticut, District of Columbia, Massachusetts, Maryland, Maine, New Jersey, New York, Pennsylvania, and Rhode Island.
   - Southeast — 119 institutions (29.1%) including the 12 states of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia.
   - Midwest — 141 institutions (34.5%) including the 13 states of Colorado, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, North Dakota, Nebraska, Ohio, South Dakota, and Wisconsin.
   - Northwest — 29 institutions (7.1%) including the 5 states of Idaho, Montana, Wyoming, Washington, and Oregon.
   - Southwest — 42 institutions (10.2%) including the 4 states of Arizona, New Mexico, Oklahoma, and Texas.
   - West — 26 institutions (6.3%) including the 3 states of California, Nevada, and Utah.

The norms sample for the *Advanced Measures of Music Audiation* was drawn to show as nearly as possible the same distribution of geographical regions, institution size within region, and public or private status as the distribution derived from the 1986-87 HEADS report.

A strictly random preliminary sample of institutions was drawn from the com-

plete list of institutions listed in the 1986-87 HEADS report. The author of the *Advanced Measures of Music Audiation* selected the group of institutions by starting at the top of the preliminary list. Using 5,000 students and 45 institutions as a base for the standardization sample, a proportional number of institutions and students was selected in each of the specified categories in terms of geographic area, size of institution, and public or private status. For the comparatively few institutions in which participation could not be arranged or in which participation did not take place, suitable substitutions were made according to the sampling specifications.

To adjust for discrepancies between the percent of students in each of the categories who were tested in the standardization program and the percent of students in each category summarized in the 1986-87 HEADS report, the test results were weighted, where appropriate, by using a weighting scale with a range from 0 through 10, in which 4 represented the base.

Description of the Norms Sample

A total of 5,336 students from 54 institutions of higher education in 27 states participated in the standardization program. Of the total sample, 3,206 were undergraduate and graduate music majors and 2,130 were undergraduate and graduate non-music majors. The test results for a total of 872 students in grades 9, 10, 11, and 12 were used to derive percentile norms for senior high schools.

Analyses of the test results in terms of mean differences indicated that there was no need to establish norms according to the geographic locations of the schools. Also, analyses of test results in terms of correlations between test scores and 1) size of the schools, 2) whether schools are public or private, 3) whether the students are undergraduates or graduates, and 4) whether the students are male or female indicated that there was no need to establish norms in accordance with those factors. Mean difference analyses did indicate, however, that it was necessary to establish separate norms for music majors and non-music majors.

No strict analysis was made to determine whether there were meaningful differences among group scores for students who are applied, music education, theory, or composition majors. The reason is that such designations are far from clear, particularly in the minds of the students themselves. It was found, nonetheless, that mean differences for groups of students in choir, band, orchestra, education classes, history classes, and theory classes were slight, and they were inconsistent from school to school.

The high school students participated in a much less extensive standardization program. They were selected on the basis of the type, location, and socio-economic status of the schools that they attended. Little attention was given to proportionality and to the representative nature of the sample.

Derivation of the Scores

The Tonal test and the Rhythm test were scored separately for each norms

group. In addition to securing raw scores for each test, the scores on the Tonal test and the Rhythm test were summed to derive raw scores for the Total test. That resulted in a set of three raw scores for each norms group: Tonal, Rhythm, and Total. Each raw score distribution was plotted and smoothed graphically, and the percentile ranks for the raw scores were read and recorded for the respective groups.

In deriving the Total score, Tonal and Rhythm raw score were not weighted through the use of standard scores. Because the differences between the standard deviations for the Tonal test score and the Rhythm test score, particularly for each norms group, are so small (the variance of a total score is not affected by the means of the tests which it comprises), and because there are only two tests to be considered, the contribution of each test score to the variance of the total test score is obvious. Therefore, it was decided that the additional step of deriving an artificially weighted total score was unnecessary. As a result, the procedure for scoring the *Advanced Measures of Music Audiation* and for deriving percentile ranks is comparatively simple.

Institutions and States Represented in the Norms Sample

University of Akron, Ohio; Anderson University, Indiana; Appalachian State University, North Carolina; University of Arizona, Arizona; Auburn University, Alabama; Bluefield College, Virginia; California State University at San Bernardino, California; Calvin College, Michigan; Capital University, Ohio; Case Western Reserve University, Ohio; Delta State University, Mississippi; Duquesne University, Pennsylvania; East Carolina University, North Carolina; Eastman School of Music, New York; Fitchburg State College, Massachusetts; Fort Hays State University, Kansas; Houghton College, New York; Hunter College, New York; University of Illinois, Illinois; Illinois Wesleyan University, Illinois; Immaculata College, Pennsylvania; Iowa State University, Iowa; Jackson State University, Mississippi; University of Kansas, Kansas; Keene State College, New Hampshire; Lebanon Valley College, Pennsylvania; University of Louisville, Kentucky; University of Lowell, Massachusetts; Loyola University, Louisiana; University of Maryland, Maryland; Memphis State University, Tennessee; Millersville University, Pennsylvania; University of Minnesota, Minnesota; Mississippi Gulf Coast Community College, Mississippi; Mississippi State University, Mississippi; University of Southern Mississippi, Mississippi; University of Northern Colorado, Colorado; Peabody Conservatory, Maryland; Pennsylvania State University, Pennsylvania; University of Pennsylvania, Pennsylvania; Potsdam State University College, New York; Southern Mississippi State University, Temple University, Pennsylvania; Texas Christian University, Texas; College of Saint Rose, New York; Samford University, Alabama; Seattle Pacific University, Washington; Shenandoah College and Conservatory, Virginia; Southeastern College, Florida; University of South Florida, Florida; University of Southern California, California; Stephen F. Austin State University, New Mexico; University of To-

ronto, Toronto, Ontario, Canada; Virginia Polytechnic Institute and State University, Virginia; and West Chester State University, Pennsylvania.

Reliability of the Tests

The stability of test scores is best interpreted through reliability information. Subtle and extraneous factors that contribute to bias and unreliability in human judgment have no effect on objective test scores. The less that test scores are influenced by a student's physical and psychological vicissitudes and by distractions within and outside the classroom, the more reliable the test. It is not possible to state categorically whether a given level of reliability associated with a test is appropriate or inappropriate. Reliability varies with the type, content, and length of a test. Reliability coefficients, however, generally range between .00 and .95. The closer the coefficient approaches the upper limit, the more stable the test scores are and the greater the freedom of the scores from factors that obscure real differences in test results. To evaluate ideally the reliability of a given test, the coefficient should be compared with those of tests that are designed for the same purpose and demonstrate similar validity.

Table 4

Advanced Measures of Music Audiation
Reliabilities, Standard Errors of Measurement, and Standard Errors of a Difference

	TONAL	RHYTHM	TOTAL	Standard Error of a Difference
		Reliabilities		
Undergraduate and Graduate Music Majors				1.7
Split-Halves	.84	.85	.88	
Retest	.86	.87	.89	
Standard Error of Measurement	1.6	1.4	2.6	
Undergraduate and Graduate Non-Music Majors				2.0
Split-Halves	.80	.80	.81	
Retest	.80	.81	.83	
Standard Error of Measurement	2.2	1.8	3.7	
High School Students				1.9
Split-Halves	.81	.82	.84	
Standard Error of Measurement	1.9	1.7	3.2	

Test reliability may be quantified in a variety of ways, which reduce to two basic types of indexes: the reliability coefficient and the standard error of measurement. The reliability coefficient is an index of the dependability of test scores for a group of students. The standard error of measurement is an index of the inconsistency of an individual student's test score.

In Table 4, two types of reliability coefficients and the standard error of measurement are reported for the *Advanced Measures of Music Audiation.* Each type of reliability is an index of the stability of test scores. The split-halves coefficient (derived from only one administration of a test) is more influenced by the homogeneity of test content. The retest coefficient (derived from two administrations of a test) is more influenced by physical and psychological changes in the student and by different environmental conditions.

The split-halves reliability coefficients are based upon the scores of the entire sample of students in each norms group that participated in the standardization program. Because of the unique procedure for scoring the *Advanced Measures of Music Audiation,* which results in adjusted scores, the procedure for deriving the split-halves coefficients also is unique. The following is a detailed description of that procedure: 1) The 20 questions on the T1 Tonal test were scored and the 20 questions on the R1 Rhythm test were scored. 2) The item difficulty and discrimination levels were computed. 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 T1 test were scored and the two halves of the T2 test were scored. 5) For each student, the score on one half of the T1 test and the score on one half of the T2 test were added together, and the score on the other half of the T1 test and the score on the other half of the T2 were added together. 6) The resulting two sets of scores were correlated and the coefficient was corrected for length through the use of the Spearman-Brown Prophecy Formula.* 7) The same procedure was followed for deriving the split-halves reliability coefficients for the Rhythm test. 8) The split-halves reliability of the Total test was derived by summing T1 and R1 and by summing T2 and R2. Those two sums were correlated and the coefficient was corrected for length through the use of the Spearman-Brown Prophecy Formula.

To determine retest reliability, a special study was undertaken. Following the standard directions, faculty members administered the *Advanced Measures of Music Audiation* to 33 undergraduate music majors and 37 undergraduate non-music majors at West Chester State University in West Chester, Pennsylvania. Exactly one week later, the test was administered again under the same conditions to those students. The two sets of scores of the music majors were correlated and the two sets of scores of the non-music were correlated.

It can be seen in Table 4 that reliability coefficients are substantial, particularly when the brevity of the complete test is considered. The differences between the split-halves and retest reliabilities are so slight that they are meaningless. That is atypical, because split-halves reliabilities are usually found to be higher than retest reliabilities for a given test. Perhaps the reason for the unusual finding has to do

with the manner in which the tests are scored and how they were divided for purposes of computing the split-halves coefficients.

Standard Error of Measurement

The standard error of measurement may be explained by an example. Consider that ten students with exactly the same arithmetic achievement all took the same arithmetic test. Regardless of their equal achievement, it would be unusual if they all received the same score on the test. Rather, their obtained scores would have a distinct range. A few students would have higher scores than they deserve, a few would have lower scores than they deserve, and the majority would have scores rather close to their "true" scores. Differences in observed and true scores are due to changing physical, psychological, and environmental conditions. The standard error of measurement is an index of the variability of the observed scores of the students who have the same true score. There is no way of determining just how much a student's true score has been underestimated or overestimated. The best estimate of a student's true score is his or her obtained score. Nonetheless, the standard error of measurement provides an interval estimate of the true score. That interval is found by adding to and subtracting from the student's obtained score the numerical value of the standard error of measurement. If a test were administered to a student 100 times, his or her true score would be included in the interval between one standard error of measurement above and below his or her obtained score approximately 66 times, the true score would be included in the interval between two standard errors of measurement above and below his or her obtained score approximately 95 times, and the true score would be included in the interval between three standard errors of measurement above and below his or her obtained score almost every time. The standard errors of measurement for the various norm groups, based upon the split-halves coefficients, for the *Advanced Measures of Music Audiation* may be found in Table 4. For making conservative interpretations, the standard error of measurement for either the Tonal test or the Rhythm test for all norm groups may be approximated as being two points.

Standard Error of a Difference

When diagnosing a student's musical strengths and weaknesses and adapting music instruction to his or her individual musical needs, idiographic evaluation is as important as normative evaluation. In idiographic evaluation, a student's score on one test is compared with his or her score on another test. The standard error of a difference is an index of the extent to which a student's scores on two tests are essentially the same or different. With regard to the *Advanced Measures of Music Audiation,* the specific interest is in whether a student's Tonal score and his or her Rhythm score are similar or different. To determine the standard error of a differ-

ence for a norms group, the variability of the tests must be comparable and the tests must have been administered to the same group or highly similar groups of students. Those conditions have been satisfied for the Tonal and Rhythm tests that constitute the *Advanced Measures of Music Audiation.* The standard error of a difference between Tonal test and Rhythm test scores for the various norm groups for the *Advanced Measures of Music Audiation* may be found in Table 4. It, like the standard error of measurement, is approximately two points.

Central Tendency and Variability of the Tests

The mean of a test, a measure of the central tendency of scores, represents the average score of all of the students in a group to whom a test is administered. It is to be expected that the mean will be different for different norm groups, depending upon the level of difficulty of the test for the given groups. The standard deviation of a test, a measure of the variability of scores, indicates the extent to which test scores vary around the mean. It serves as a reference for determining how much higher and how much lower some scores are than others as compared with the average score. Given a large enough group of students and an appropriate test, it may be assumed that scores will be distributed up to three standard deviations above the mean and up to three standard deviations below the mean. Such a distribution is referred to as a "normal curve." Approximately thirty-four percent of the scores will fall somewhere between the mean and one standard deviation above the mean. Approximately fourteen percent of the scores will fall somewhere between the first and second standard deviations above the mean. Approximately two percent of the scores will fall somewhere between the second and third standard deviations above the mean. The approximate percentages are the same for scores below the mean. When the standard deviation is used as a guide, it is evident that the majority of students, sixty-eight percent, are considered average, fourteen percent are considered above average, and fourteen percent are considered below average. Two percent are considered exceptionally good and two percent are considered exceptionally poor.

The means and standard deviations for the various norm groups for the *Advanced Measures of Music Audiation* may be found in Table 5. The relationship of the mean and standard deviation of each test suggests that the ranges between the average score and the highest score and between the average score and a "chance" score conform quite closely to a normal distribution. That indicates that the overall difficulty of the tests is appropriate. It is interesting to note that the means and standard deviations of the tests are rather similar for the non-music major undergraduate and graduate college and university students and the high school students. As can be seen in Tables 1, 2, and 3, the raw score-percentile rank equivalents are exactly the same in the lower half of the distribution for those two groups of students. It has been found that chronological age has very little effect on test results, and because both groups included some students with a musical back-

ground, even though they were not music majors, members of school music ensembles, or currently taking private lessons, the results should not be surprising.

Table 5

Advanced Measures of Music Audiation
Means and Standard Deviations

	Mean	Standard Deviation
Undergraduate and Graduate Music Majors		
TONAL	28.3	4.12
RHYTHM	30.8	3.52
TOTAL	59.1	7.41
Undergraduate and Graduate Non-Music Majors		
TONAL	24.3	4.89
RHYTHM	27.4	4.11
TOTAL	51.7	8.49
High School Students		
TONAL	23.8	4.37
RHYTHM	26.8	4.03
TOTAL	50.6	7.91

Table 6

Advanced Measures of Music Audiation
Intercorrelations Among Tests

Undergraduate and Graduate Music Majors

	Tonal	Rhythm	Total
Tonal	---	.78	.94
Rhythm	.78	---	.93
Total	.94	.93	---

Undergraduate and Graduate Non-Music Majors

	Tonal	Rhythm	Total
Tonal	---	.72	.93
Rhythm	.72	---	.91
Total	.93	.91	---

High School Students

	Tonal	Rhythm	Total
Tonal	---	.74	.95
Rhythm	.74	---	.94
Total	.95	.94	---

Intercorrelation of Test Scores

The extent to which the Tonal test and the Rhythm test of the *Advanced Measures of Music Audiation* measure unique dimensions of music aptitude is determined through the examination of intercorrelation coefficients. The Tonal test scores and Rhythm test scores intercorrelation and the intercorrelations of those scores with the Total test score for the various norm groups are reported in Table 6. The intercorrelations between the Tonal test scores and the Rhythm test scores are unusually high when compared with intercorrelations found among tests of a similar type. That expected undesirable finding is in great part a result of the fact that the ten questions which have 'Same' as the correct answer constitute half of the number of questions on the Tonal test and half the number of questions on the Rhythm test. As high as the intercorrelation coefficients are, they are nevertheless lower than the reliabilities of either of the tests. Thus it is clear that although the Tonal test and the Rhythm test have much in common, there is still a substantial portion of the variance of each test that is unique in terms of the other.

Item Difficulty and Discrimination

The test questions included in the Tonal test and the Rhythm test of the *Advanced Measures of Music Audiation* were selected according to two criteria: difficulty levels and point bi-serial item-test discrimination values. As can be seen in Table 7 (for ease of reading, decimals are omitted from the table), there is a considerable range of difficulty among the questions. That is as it should be. (The higher the difficulty value, the easier the question.) Moreover, all of the question have discrimination values of .20 and higher, many of them being well above the minimum standard of .20. The average item difficulty level was found to be 67.7. The average discrimination level was found to be 40.2. The item difficulty and discrimination indexes were derived from the scores of the entire sample of students, undergraduates and graduates and music majors and non-music majors, that participated in the standardization program. Scores on the comprehensive test of thirty test questions, questions which were not separated into a Tonal test and a Rhythm test, served as the basis for computing the item characteristics.

Validity of the Tests

The Nature of Validity

The types of statistical data that might be investigated as evidence of the validity of a test are means, standard deviations, reliability coefficients, test intercorrelations, standard errors of a difference, difficulty indexes of test items, discrimination values of test items, item homogeneity within each subtest, and correlations of

test scores with other measures such as course grades, teachers' evaluations, scores on similar tests, performance achievement, later success in music, and experimental measures of essentially the same type of behaviors that the test is designed to measure. All of those data yield useful information for evaluating the validity of a test. Nonetheless, there is no absolute index of the validity of a test. The presence or absence of certain qualities of a test does or does not prove that a test measures what it "purports" to measure. Moreover, data cannot prove that the traits measured in a test are those that should be measured. It is impossible to prove that a test measures what it is designed to measure. Thus no test has perfect validity. A test may be more valid for one purpose and less valid for another. For those reasons, among others, subjective considerations as well as objective evidence are important in assessing the validity of a test. Both subjective considerations and

Table 7

Advanced Measures of Music Audiation
Item Difficulty and Discrimination Indexes

Item Number	Difficulty Level	Discrimination Value
1	59	49
2	66	63
3	97	24
4	59	47
5	73	36
6	55	32
7	60	55
8	70	53
9	84	38
10	93	25
11	94	43
12	87	30
13	82	33
14	59	47
15	85	38
16	92	35
17	99	32
18	27	20
19	74	60
20	87	30
21	90	40
22	55	20
23	59	35
24	27	42
25	36	36
26	50	69
27	76	32
28	33	30
29	57	47
30	47	64

objective evidence are offered in support of the validity of the *Advanced Measures of Music Audiation.*

Subjective Validity

Content validity and construct validity are the two most important types of subjective validity, perhaps even the two most important types of overall test validity. Unless a teacher who uses a test is satisfied that the content of the test fairly represents factors which should be measured (content validity) and that the manner in which the test content is measured is appropriate (construct validity), he or she will have no confidence in the results of the test even if the objective validity of the test is overwhelming.

As a preliminary step in assessing the content and construct validity of the *Advanced Measures of Music Audiation,* Part 2 and Part 3 of this manual should be reviewed. To assess how well the content and design of the *Advanced Measures of Music Audiation* conform to what is known about music aptitude and audiation, knowledge of the rationale of the *Advanced Measures of Music Audiation* in relation to the nature, characteristics, and description of music aptitude is necessary. Moreover, an understanding of the role of audiation in music aptitude and musicianship is essential. Detailed information about music aptitude and audiation beyond that offered in this manual may be found in selected publications listed in the bibliography.

Objective Validity

In order to carry out an objective validity study of a test, at least one valid criterion measure with which to correlate test scores must be identified. That is not an easy task when one is attempting to validate a test designed for college and university music majors. The problem is that professors of music show very little variation in the grades that they give, most of the grades being A or B. That is particularly true for grades awarded in applied music and performance organizations. The magnitude of a correlation coefficient depends upon at least reasonable variability in the two factors being correlated, the validity criterion and the test itself; therefore objective validity for tests designed for use with college and university music majors is difficult to establish. It has been found by the writer in several unpublished studies, using the *Musical Aptitude Profile* and nationally recognized academic (non-music) achievement tests, that grades in music correlate significantly higher with the scores students attain on the non-music academic achievement tests than with those on the music aptitude test. That is surprising because, although the *Musical Aptitude Profile* was not meant to be used with musically select students beyond grade twelve, the battery has been shown to possess extra-ordinarily high experimental validity when used with students in elementary through secondary schools. Because the academic tests include no listening to music and no questions about music, and because it is realistic to expect that the *Musical Aptitude Profile* should demonstrate at least some correla-

tion with music behaviors, even when used with older students, it is reasonable to conclude that many music professors grade more on students' academic intellectual achievement than on the students' musicianship, or even their knowledge about music. That is to say, the validity criteria traditionally used are themselves invalid.

Until valid criterion measures can be established, it is not possible to investigate the criterion-related (objective) validity of the *Advanced Measures of Music Audiation.* Studies to achieve that goal and to establish longitudinal predictive validity for the *Advanced Measures of Music Audiation* are now in process. In addition, the effects of practice and training on *Advanced Measures of Music Audiation* scores are being investigated. The results will be published as soon as the studies are completed. In place of the traditional criterion-related validity studies, it is hoped that the reader will accept the suppositions and information presented in the remainder of this part of the manual as at least tentative objective validity of the *Advanced Measures of Music Audiation.*

It was found in the standardization program that the difference between the scores of graduate students and undergraduate students, both music majors and non-music majors, is insignificant. That finding supports the belief that the *Advanced Measures of Music Audiation* is a music aptitude test and not a music achievement test. If it were a music achievement test, it would be expected that older students with a broader music education would score much higher on the test than students with a less extensive music background. Moreover, the average scores of college and university non-music majors, both undergraduate and graduate, and those of high school students are not materially different. Again, chronological age, experience, and music education appear to have little relationship to scores on the *Advanced Measures of Music Audiation.*

A study to investigate further the effect of chronological age on *Advanced Measures of Music Audiation* scores was undertaken early in 1989 in Northfield, Minnesota. The test was administered by the conductor to members of the Northfield Children's Chorus, a musically select group of 61 students in grades 4 through 7, ages 9 through 12. The scores ranged from 18 to 32 on the Tonal test, from 15 through 34 on the Rhythm test, and from 34 to 66 on the Total test. Of most significance are the measures of central tendency and variability. The means and standard deviations were 24.6 and 4.09 for the Tonal test, 27.1 and 4.21 for the Rhythm test, and 51.9 and 7.83 for the Total test. As can be seen in Table 5, those data are almost identical with those found for college and university non-music majors in the standardization program.

It may seem logical to assume that because college and university music majors score significantly higher than non-music majors on the *Advanced Measures of Music Audiation,* the test must be a music achievement test. Such differences between music and non-music students, however, have consistently been found over the years when objectively and longitudinally validated music aptitude tests designed for younger students have been used. The reason seems to be that a self selective process is operating. It is natural for some students with high levels of

music aptitude to follow music both educationally and vocationally, whereas students with low levels of music aptitude tend less to do so. Historically, that music students score higher than non-music students and that practice and training have a negligible effect on test scores have continually supported the validity of the *Musical Aptitude Profile.*

It was also found in the standardization program that *Advanced Measures of Music Audiation* score distributions were virtually the same for college and university students regardless of the location of the school that they attended, the size of the school that they attended, or the type of the school that they attended. Those facts bode well for the validity of the *Advanced Measures of Music Audiation* as a music aptitude test. If the *Advanced Measures of Music Audiation* was actually a music achievement test, the distributions of scores of students attending institutions in different areas of the country would be dissimilar, as would the distributions of scores of students attending institutions of different sizes and the distributions of scores of students attending public and private institutions. Of course, such differences in the score distributions would not be perfectly consistent.

A special type of objective validity is congruent validity. To establish the validity of one test, students' scores on that test are correlated with their scores on another test that is designed to measure the same factor, and that has established objective validity. To gain insight into the validity of the *Advanced Measures of Music Audiation,* scores on that test were correlated with scores on the *Musical Aptitude Profile.* However, the results cannot be interpreted strictly as congruent validity for three reasons. First, the *Musical Aptitude Profile* was not designed to be used with college and university students. Second, the validity of the *Musical Aptitude Profile* when used with college and university students has not been established. Third, because the *Musical Aptitude Profile* and the *Advanced Measures of Music Audiation* have different item difficulty characteristics, the score distributions for the two tests are dissimilar. It is for the third reason particularly that the correlation coefficients which relate to congruent validity would not be expected to be exceptionally high. Still, because the *Musical Aptitude Profile* has the most varied and the highest validity of published music aptitude tests that can be used with college and university students, and because the data were readily available in two schools, the analyses were undertaken. It was reasoned that if even moderate relationships between scores on the two tests were found, that would at least indirectly give support to the validity of the *Advanced Measures of Music Audiation* as well as offer positive direction for designing more pertinent validity studies of the test. Students who took part in three congruent validity studies were enrolled in the University of Louisville, Texas Christian University, and Illinois Wesleyan University.

The complete *Musical Aptitude Profile* was administered to 20 undergraduate music majors at the University of Louisville approximately one year before the *Advanced Measures of Music Audiation* was administered to them in the standardization program. The correlations are reported in Table 8.

Table 8

Correlations Between the Advanced Measures of Music Audiation and the Complete Musical Aptitude Profile

	Advanced Measures of Music Audiation		
	Tonal	Rhythm	Total
Musical Aptitude Profile			
Tonal Imagery-Total	.73	.70	.74
Rhythm Imagery-Total	.67	.71	.72
Musical Sensitivity-Total	.46	.43	.57
Composite	.71	.76	.78

The correlations were found to be much higher than anticipated, particularly when it is remembered that the sample was small and homogeneous and that the two tests were not administered concurrently, an important condition for establishing congruent validity. Another encouraging result is that the scores on the *Musical Sensitivity-Total* test, a preference test, correlate least with all of the scores on the *Advanced Measures of Music Audiation,* which are non-preference tests.

Only the four non-preference subtests of the seven subtests of the *Musical Aptitude Profile* were administered to 33 undergraduate music majors at Texas Christian University approximately one year before the *Advanced Measures of Music Audiation* was administered to them in the standardization program. The four subtests were the *Tonal Imagery-Melody* test, the *Tonal Imagery-Harmony* test, the *Rhythm Imagery-Tempo* test, and the *Rhythm Imagery-Meter* test. The correlations are reported in Table 9.

Table 9

Correlations Between the Advanced Measures of Music Audiation and Selected Subtests of the Musical Aptitude Profile

	Advanced Measures of Music Audiation		
	Tonal	Rhythm	Total
Musical Aptitude Profile			
Tonal Imagery-Melody	.76	.64	.81
Tonal Imagery-Harmony	.73	.59	.77
Rhythm Imagery-Tempo	.46	.70	.68
Rhythm Imagery-Meter	.60	.70	.85

Although they are of like magnitude, the results of the Texas study are even more compelling than those of the Kentucky study in that they are more clear cut. It can be seen in Table 9 that the two tonal subtests of the *Musical Aptitude Profile* are more highly correlated with the Tonal test of the *Advanced Measures of Music Audiation* than they are with the Rhythm test of the *Advanced Measures of Music Audiation.* Similarly, the two rhythm subtests of the *Musical Aptitude Profile* are

more highly correlated with the Rhythm test of the *Advanced Measures of Music Audiation* than they are with the Tonal test of the *Advanced Measures of Music Audiation.*

Only the Composite score on the Musical Aptitude Profile was available for 37 undergraduate music majors at Illinois Wesleyan University who were given the test approximately one year before they took the *Advanced Measures of Music Audiation* in the standardization program. The correlation between the Composite score on the *Musical Aptitude Profile* and the Total score on the *Advanced Measures of Music Audiation* was .58. There appears to be no ready explanation for such a low coefficient. However, as low as it is when compared with the coefficients found in the previous two studies, it nonetheless indicates that there is substantially more than a chance relationship between scores on the two tests.

Validity of the Adjusted Scores

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. As a matter of fact, 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 on the *Advanced Measures of Music Audiation* are adjusted on the basis of how well a student positively and negatively audiates tonally and how well a student positively and negatively audiates 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 to subtract a point because a student believed that 1) a musical answer is different from a musical statement as a result of a rhythm change when in fact the musical answer sounds the same as the musical question, or the musical answer sounds different as a result of a tonal change, or 2) a musical answer is different from a musical statement as a result of a tonal change when in fact the musical answer sounds the same as the musical question, or the musical answer sounds different as a result of a rhythm change.

Subjective validity, in terms of both content and construct validity, is enhanced by the adjustment of scores on the *Advanced Measures of Music Audiation.* Moreover, there is preliminary objective evidence to suggest that the adjusted scores are more accurate than the unadjusted scores. The results of the testing of the total sample of 5,336 students who participated in the national standardization program of the *Advanced Measures of Music Audiation* were analyzed to investigate the relationship among the unadjusted and adjusted test scores. That information in the form of correlation coefficients is presented in Table 11. The means and standard deviations of the respective tests are presented in Table 10.

Table 10

Means and Standard Deviations for the Unadjusted and Adjusted Test Scores for the Complete Standardization Group

	Mean	Standard Deviation
Unadjusted Tonal Test	11.3	3.39
Unadjusted Rhythm Test	11.8	2.90
Unadjusted Total Test	23.1	5.80
Adjusted Tonal Test	26.3	4.78
Adjusted Rhythm Test	28.9	4.15
Adjusted Total Test	55.2	8.54

Table 11

Correlations Among the Unadjusted and Adjusted Tests Scores for the Complete Standardization Group

	Adjusted Scores		
	Tonal	Rhythm	Total
Unadjusted Scores			
Tonal	.89	.81	.91
Rhythm	.80	.87	.90
Composite	.88	.86	.93

It can be seen in Table 11 that the unadjusted and adjusted Tonal test scores correlate .89, the unadjusted and adjusted Rhythm test scores correlate .87, and the unadjusted and adjusted Total test scores correlate .93. As should be expected, particularly with such a large group of heterogeneous music majors and non-music majors, the correlations are high. They are not, however, perfect, as would be anticipated if the adjusted scores were traditional corrected-for-guessing scores. The amount of unique variance associated with each of the adjusted scores offers evidence that the adjusting of scores on the *Advanced Measures of Music Audiation* is justified. Moreover, the reliabilities of the unadjusted scores are lower than those for the adjusted scores. For the heterogeneous group of all students who participated in the standardization program, the reliabilities of the adjusted scores are .89 for the Tonal test, .90 for the Rhythm test, and .92 for the Total test. The reliabilities of the unadjusted scores are .76 for the Tonal test, .76 for the Rhythm test, and .80 for the Total test. Those coefficients are indicative of the negative correlations of $-.51$ and $-.49$ between the number of points subtracted from the actual scores and the adjusted scores on the Tonal test and the Rhythm test. Conclusive evidence of the value of the adjusted scores will be a part of the results of the longitudinal predictive validity studies of the *Advanced Measures of Music Audiation* that are now taking place.

* Questions 2, 7, 9, 11, 12, 14, 18, 25, 27, and 29 were included in one half of the T1 test and questions 1, 5, 6, 13, 16, 19, 20, 22, 23, and 24 were included in the other half of the T1 test. Questions 3, 6, 8, 15, 17, 18, 20, 22, 27, and 29 were included in one half of the R1 test and questions 2, 4, 9, 10, 13, 21, 24, 26, 28, and 30 were included in the other half of the R2 test.