A multi-phase examination of Advanced Measures of Music Audiation at three American music schools

David A. Stringham, James Madison University
A. H. Snell, II
R. F. Grunow
A Multi-Phase Examination of Advanced Measures of Music Audiation at Three American Schools of Music

David A. Stringham  
James Madison University

Alden H. Snell II  
University of Delaware

Richard F. Grunow  
Eastman School of Music

Abstract

In the early- and mid-20th century, numerous researchers developed tests of music ability, aptitude, intelligence, musicality, and talent. These persons and their research influenced Gordon, who remains the foremost developer of music aptitude measures. In this report, we describe results from research with Gordon’s Advanced Measures of Music Audiation (AMMA).

We examined predictive validity of AMMA, and differences in AMMA scores by instrument group in a three-part study of undergraduate students. We conducted (a) preliminary inquiries using subsets of data from one institution, (b) a pilot study with the entire sample from one institution, and (c) an expanded study using data from three institutions (N = 372) of varying size and selectivity.

Findings offer preliminary evidence to suggest that (a) AMMA is a reliable measure of composite stabilized music aptitude and (b) statistically significant differences in music aptitude scores by instrument group are not consistent among institutions. Future researchers might gather and analyze similar data from other institutions to examine whether published norms for AMMA remain representative of the population for which they are intended.

Keywords

music aptitude, Advanced Measures of Music Audiation, higher education, music instrument comparisons, predictive validity

In the early- and mid-20th century, numerous researchers developed tests of music ability, aptitude, intelligence, musicality, and talent. Schoen’s “Tests of Musical Feeling and Musical Understanding” (1925) was the first of such instruments, followed by those of Bentley (1966), Drake (1954), Gaston (1957), Kwalwasser (1953), Kwalwasser and Dykema (1930), Lowery (1926), Lundin (1949), Madison (1942), Orman (n.d.), and Wing (1958). Seashore’s Measures of Musical Talent (1919), subsequently revised in collaboration with Lewis and Saetveit in 1960, was examined in a number of studies (e.g., Seashore, 1938; Stanton, 1935); this instrument was perhaps most significant of these. The aforementioned researchers influenced Gordon, who defined music aptitude as an individual’s “potential to learn music” (Gordon, 2012, p. 44). While other researchers have documented expanded discussion
regarding music aptitude (e.g., Grashel, 2008; Haroutounian, 2000), Gordon (1965, 1979, 1982, 1988, 1989a, 1989b, 1995) has remained the foremost developer of music aptitude measures, examined in a variety of studies (e.g., Gordon, 1967; Schleuter, 1974, 1977; Snell, 2005; Stamou, Schmidt, & Humphreys, 2010).

One of Gordon’s most recently-developed measures, Advanced Measures of Music Audiation (AMMA; Gordon, 1989a), is a measure of stabilized music aptitude; Gordon reports that a person transitions from developmental to stabilized music aptitude at approximately age nine (Gordon, 2012, p. 45), and that music aptitude is normally distributed among a population (Gordon, 1987, p. 12). AMMA comprises thirty items; for each item, students listen to a short musical statement and short musical answer, and are asked to report whether they hear a tonal change, a rhythm change, or no change between the statements. Gordon developed AMMA at the request of music administrators (1997, p. 72); AMMA is intended to be predictive of music achievement, and to measure music aptitude regardless of music achievement (e.g., ability to perform vocally or instrumentally, read music, or understand music theory). Researchers have examined reliability and validity of AMMA (e.g., Estrella, 1992; Fullen, 1993; Gordon, 1989c, 1990, 1991a, 1991b, 1997; McCrystal, 1995; Stringham, 2010); however, relatively few studies document descriptive statistics, reliability, or validity of AMMA in college settings for which AMMA was developed.

Purpose and Need for the Study

In this report, we summarize findings from an evolving research project examining administrations of AMMA at three American music schools. With the purpose of improving music teaching and learning in collegiate settings by enhancing our profession’s understanding of AMMA, we sought to (a) examine the extent to which AMMA predicted academic outcomes (e.g., acceptance, graduation, GPA), (b) document descriptive statistics from multiple administrations of AMMA at American music schools, and (c) compare raw scores based on major, institution, and instrument group.

We designed this study to explore three gaps in literature: (1) Few descriptive statistics for administrations of AMMA among college students are available (Estrella, 1992; Gordon, 1989c, 1990, 1991b, 1997; McCrystal, 1995); fifteen years after the most recent of these studies, we were curious whether descriptive statistics have changed substantially. (2) Gordon standardized AMMA more than twenty years ago (1989a); we were curious whether descriptive statistics among undergraduate students had changed considerably. (3) Findings of differences in AMMA scores by instrument group (Gordon, 1989c; Snell, 2010) warranted further exploration. In the test manual, Gordon reports:

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 were inconsistent from school to school (1989c, p. 38).

Method

Our inquiry originated in administration of AMMA at a school of music in the northeastern United States. From 2003–2006, a faculty member at this institution administered AMMA to auditioning undergraduate students; since 2007, the same faculty member has administered AMMA to matriculated
undergraduate students. The research project documented here occurred in three phases: (a) two preliminary inquiries conducted with a portion of this sample data, (b) a pilot study using the complete data sample from this institution, and (c) an expanded study using data from three higher education institutions.

**Preliminary Inquiries**

In an initial analysis of data from 2008, Stringham (2009) revealed predictive validity for an offer of admission to this institution; however, in this sample, AMMA was not predictive of enrollment at this institution or first-semester academic achievement. Snell (2010) analyzed data from students who enrolled between 2003 and 2009, to determine whether statistically significant differences in AMMA scores were found based on instrument group. In this sample, keyboard (i.e., piano and organ) majors’ AMMA scores were consistently highest; vocalists’ scores were consistently lowest. Based on these results, we designed a pilot study, in which we examined whether trends found in these preliminary inquiries were present in the complete sample.

**Pilot Study**

Using data gathered in administrations of AMMA and from the institution’s registrar, our pilot study included the entire sample (N = 938) of de-identified secondary data from students who enrolled at this institution (Stringham, Snell, & Grunow, 2010). We asked the following research questions: (1) How do descriptive statistics from these administrations of AMMA compare with others in literature? (2) Do AMMA scores differ by (a) major, (b) instrument group, or (c) year? and (3) Do scores on AMMA predict: (a) first-semester grade point average, (b) cumulative grade point average, and (c) graduation from this institution?

Composite split halves reliability for AMMA in this sample was acceptable (r = .77). Results revealed that: (1) In comparison with published norms, mean scores were higher and variability was lower (see Table 1); and (2) Scores differed by instrument group and year, and noticeably by major. (Table 2 is on page 13.)

<table>
<thead>
<tr>
<th>Table 1. AMMA Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Sample Composite</td>
</tr>
<tr>
<td>AMMA Norms</td>
</tr>
</tbody>
</table>

Note: **TM** = tonal mean, **TSD** = tonal standard deviation, **RM** = rhythm mean, **RSD** = rhythm standard deviation, **CM** = composite mean, **CSD** = composite standard deviation.

Initially, we planned to conduct linear regression analysis to answer the first two parts of our third research question (relationships between stabilized music aptitude and (a) first-semester grade point average and (b) cumulative grade point average). To examine the third part of this question (relationship between stabilized music aptitude and graduation), we planned to conduct logistic regression analysis. Prior to conducting these analyses, we examined (a) correlations between stabilized music aptitude and each variable, and (b) each variable’s distribution to ensure that these analyses were appropriate.
After examining the relationship between music aptitude and grade point average (first semester and cumulative), we determined that linear regression analysis would not be appropriate because: (a) music aptitude and the two outcome variables correlated weakly, (b) grade point average distributions were leptokurtic and exhibited considerable negative skewness, and (c) mean grade point average was higher, and variability lower, than suggested theoretical mean and standard deviation.

Finally, to examine the relationship between stabilized music aptitude and graduation, we conducted logistic regression. This procedure revealed that stabilized music aptitude, measured using AMMA, was not a significant predictor of graduation \( \chi^2 (1, N = 933) = 2.57, p = 0.11 \). Classification was unimpressive, with a success rate of 54.7 percent.

The most salient finding from this study appears to be considerable differences in AMMA scores by instrument group, confirming findings from previous research (Snell, 2010). We hypothesized that the lack of predictive validity in this inquiry may have been due to homogeneity of the sample. In the third phase of our study, we chose to focus on differences in AMMA scores by instrument group, but further expand our research to three institutions.

**Current Study**

In Fall 2010, we expanded our study of AMMA (Gordon, 1989a) to examine data gathered at three American schools of music. To do so, we retained data from our pilot study, and added data from two other institutions. Our research questions were: (1) How do descriptive statistics from administration of AMMA at three institutions compare with each other and with published norms? and (2) Do AMMA scores differ by instrument group at these three institutions?

**Research setting, subjects, and Research Subjects Review Board.** In addition to pilot study data from a medium, private research university in the northeastern United States (hereafter, School 1), we gathered data from schools of music at two other institutions: (a) School 2, a large public university in the southeastern United States (n = 98), and (b) School 3, a private liberal arts college in the northeastern United States (n = 57). Because we had different amounts of data from each institution (eight years from School 1, one year from School 2, and two years from School 3), we excluded data from School 1 gathered prior to 2009. Data from 372 students were in the composite sample. Each institution approved research protocol for this study. Study data was stored on a password-protected hard drive in a locked office, accessible only to the researchers.

**Data sources, collection, and analysis.** AMMA is administered in approximately 20 minutes. Research suggests it is a reliable instrument for measurement of music aptitude (e.g., Gordon, 1989c, 1990, 1991a, 1991b; McCrystal, 1995; Stringham, 2010).

To ensure that AMMA was a reliable instrument for this sample, we calculated split-halves reliability for the administration of AMMA at each of these institutions. For our sample, reliability coefficients for tonal and rhythm dimensions were .52 and .57, respectively; our composite scores resulted in a reliability coefficient of .73. Specific findings for each school, along with those from our sample and AMMA norms, are in Table 2.
Table 2. Comparison of Sample and National Split-Halves reliability for AMMA

<table>
<thead>
<tr>
<th></th>
<th>Tonal</th>
<th>Rhythm</th>
<th>Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>0.53</td>
<td>0.46</td>
<td>0.70</td>
</tr>
<tr>
<td>School 2</td>
<td>0.54</td>
<td>0.37</td>
<td>0.64</td>
</tr>
<tr>
<td>School 3</td>
<td>0.60</td>
<td>0.64</td>
<td>0.82</td>
</tr>
<tr>
<td>Sample Composite</td>
<td>0.57</td>
<td>0.52</td>
<td>0.73</td>
</tr>
<tr>
<td>AMMA Norms</td>
<td>0.84</td>
<td>0.88</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Note: Sample reliability corrected for length using Spearman-Brown Prophecy Formula. School 1, N = 217; School 2, N = 98; School 3, N = 57. Sample Composite, N = 372; AMMA Norms, N = 3,206.

Officials at each institution provided participants’ tonal, rhythm, and composite scores on AMMA, and major instrument. After receiving this de-identified data, we coded variables to represent each student’s major instrument. To ensure accuracy, two researchers independently coded data and compared results. Data were imported into PASW Statistics (Version 17) for further analysis.

Results

To answer the first research question, “How do descriptive statistics from administration of AMMA at three institutions compare with each other and with published norms?,” we calculated descriptive statistics for each institution, as well as the total sample. For the tonal dimension, scores at all three schools were higher than those in published norms. For the rhythm dimension, only scores at School 1 were higher than those in published norms. For the composite dimension, scores at Schools 1 and 2 were higher than those in published norms. Complete descriptive statistics from our sample, along with published norms (Gordon, 1989c), are in Table 3.

Table 3. AMMA Descriptive Statistics by Institution

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>TM</th>
<th>TSD</th>
<th>RM</th>
<th>RSD</th>
<th>CM</th>
<th>CSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>School 1</td>
<td>217</td>
<td>30.08</td>
<td>4.05</td>
<td>31.89</td>
<td>3.33</td>
<td>62.00</td>
<td>7.11</td>
</tr>
<tr>
<td>School 2</td>
<td>98</td>
<td>28.64</td>
<td>3.72</td>
<td>30.74</td>
<td>3.26</td>
<td>59.33</td>
<td>6.68</td>
</tr>
<tr>
<td>School 3</td>
<td>57</td>
<td>27.39</td>
<td>4.27</td>
<td>29.30</td>
<td>3.72</td>
<td>56.67</td>
<td>7.43</td>
</tr>
<tr>
<td>Sample Composite</td>
<td>372</td>
<td>29.29</td>
<td>4.11</td>
<td>31.19</td>
<td>3.49</td>
<td>60.48</td>
<td>7.31</td>
</tr>
<tr>
<td>AMMA Norms</td>
<td>3,206</td>
<td>23.80</td>
<td>4.37</td>
<td>30.80</td>
<td>3.52</td>
<td>59.10</td>
<td>7.41</td>
</tr>
</tbody>
</table>

Note: TM = tonal mean, TSD = tonal standard deviation, RM = rhythm mean, RSD = rhythm standard deviation, CM = composite mean, CSD = composite standard deviation.

To answer the second research question, “Do AMMA scores differ by instrument group at these three institutions?,” we calculated descriptive statistics for AMMA by institution, instrument group, and year. AMMA results were consistent by year; therefore, we are presenting descriptive statistics by institution
and instrument group. Compared to published norms, mean scores were higher and variability was lower at each institution. Generally, mean AMMA scores were highest at School 1 and lowest at School 3; however, congruent with Gordon (1989c) and Snell (2010), there are inconsistencies in mean AMMA scores among instrument groups. At School 1, statistically significant differences occur between instrument groups on stabilized tonal, rhythm, and composite music aptitude. At School 2, statistically significant differences occur between instrument groups on stabilized tonal music aptitude. No statistically significant differences were observed at School 3. Data are displayed in Figure 1 (mean tonal AMMA results by institution and instrument group), Figure 2 (mean rhythm AMMA results by institution and instrument group), and Figure 3 (mean composite AMMA results by institution and instrument group).

![Figure 1. Mean Tonal AMMA Results by Institution and Instrument Group](image1)

![Figure 2. Mean Rhythm AMMA Results by Institution and Instrument Group](image2)
Conclusions and Future Research

This study offers preliminary evidence to suggest that (a) AMMA is a reliable measure of composite stabilized music aptitude; and (b) statistically significant differences in music aptitude scores by instrument group are not consistent among institutions. Differences observed in mean and standard deviation between institutions may be explained by selectivity and homogeneity of each school’s music major population. Of the three institutions from which we gathered data for this study, School 1 is the most selective and most homogeneous; School 3 is the least selective and most heterogeneous. Generalization of these findings is not recommended, pending further research with this sample and others.

A logical next step in our research is gathering and analyzing data from other institutions, similar to and different from those represented in this study. Additional data will help us understand if findings from this study persist at other institutions of varying size, location, and selectivity. For example, are there statistically significant differences between instrument groups at other highly selective private schools of music? Are students at small liberal arts colleges more heterogeneous in terms of music aptitude? Future researchers, when considering data gathered from a variety of institutions, should examine whether published norms for this test remain representative of the population for which they are intended.

References


**Bios**

**David A. Stringham** is Assistant Professor of Music at James Madison University. His research interests include creativity and music teacher education.

**Alden H. Snell II** is Instructor of Music Education at the University of Delaware. His research interests include teacher musicianship and professional development.

**Richard F. Grunow** is Professor of Music Education at the Eastman School of Music. His research interests include applications of Music Learning Theory to instrumental music instruction and measurement and evaluation of musical behavior.