The Primary Measures of Music Audiation (PMMA) test, published by GIA Publications, Inc., was designed by Edwin Gordon in 1979. The test purports to measure the musical aptitude of children from kindergarten through the third grade. The test features 2 audiotapes that test the student on two basic musical parameters: pitch and rhythm discrimination. The student hears two musical examples in rapid succession, then is asked if the examples are the same or different. (Gordon, 1972) The manual provides data on reliability and validity, and separate studies have examined the reliability and validity as they pertain to students of other cultures and mental statuses.

The problem, as Gordon (1972) frames it, is that tests of musical aptitude given to children before age 9 were displaying inconsistent results. He believes this to be because musical aptitude is a fluid trait that develops over the course of a child’s life, stabilizing around the age of 9. He set out to design a test that was accessible to students of such a young age that would yield reliable results that were not conflagrated with tests of abstract pitch/rhythmic memory.

This problem of conflagration led him to develop the idea of “audiation,” which is the act of hearing music “through recall or creativity.” (Gordon, 1972, p. 43) That is, the students must keep the sound of two different audio samples in their memory, accessing the possible similarities/differences without physically hearing the examples being played at the moment of decision-making. For the purposes of this test, these audio clips must be musical, as opposed to discrete abstract sounds, because students might have other
perceptual abilities, harmonic or rhythmic, which they could bring to bear in answering the items. (Gordon, 1972) This would be a mediating variable that would skew the data in favor of a student who may have aural aptitude, but not ‘musical’ aptitude.

The test consists of two twenty-minute audiotapes, each with twelve total minutes of audio test items. The student listens to two items in rapid succession, then circles one of two answers: two matching images if the clips were the same or two different images if the clips were not the same. The clips are standardized for each grade level, developed by Gordon based on his own research of children’s perceptual abilities. The test results in three scores: tonal, rhythm, and composite. (Gordon, 1972)

The standardization sample for the test consisted of 873 students from kindergarten through fourth grade in upper New York. (Gordon, 1972) Some deem this as being too small of a normalization sample. (Lehman; Wehner, 1985) The manual claims that the sample was representative of average elementary school students in terms of gender, race, socio-economic status, IQ, etc. (Wehner, 1985) Without obtaining the manual, specific numbers can’t be given to this claim.

The manual reports split-test reliability across the four grade levels as .85 to .89 for the tonal scores, .72 to .86 for the rhythm scores, and .90 to .92 for the composite scores. The standard error of measurement was reported as ranging from 1.9 to 2.5. (Lehman; Wehner, 1985) Additional experiments, also performed by Gordon, on children from a community music school and a private academic institution in western New York yielded split-half reliability scores of .70 to .91 for the tonal scores and .68 to .81 for rhythm scores. (Gordon, 1972)
The manual also reports values for test-retest reliability as .68 to .73 for the tonal scores, .60 to .73 for the rhythm portion, and .73 to .76 for the composite scores. The standard error of measurement was reported here as ranging from 1.3 to 2.0. (Lehman; Wehner, 1985)

As a starting point for considerations of validity, this test does appear to test the construct of musical aptitude in children. It gives a distinct operational definition of musical aptitude, and presents a battery of items that appear to be measuring a child’s cognitive musical functioning. Thus, it would be deemed to have face validity. While not a quantifiable form of validity, it does instill confidence in a researcher that more rigorous investigations of validity will prove fruitful.

In his explication of “audiation,” Gordon conceptually defines the content validity of the PMMA. He is careful to discriminate between musical achievement and musical aptitude, so that the test scores can be used to identify students’ possible musical abilities whether or not they have received training. He is also careful to define musical aptitude as a fluid construct for young children, which sets this test apart from others. The procedure remains the same across both subgroups of aptitudes, tonal and rhythmic; no musical or even lingual reading skills are necessary. (Gordon, 1972) The audio examples are placed in musical contexts, rather than being presented as abstract sounds. These steps appear to control for mediating or third variables such as confusion regarding instructions, different reading levels, or alternative perceptual memory modalities. (Gordon, 1972)

In 1983, Louis C. Woodruff published his dissertation, “A Predictive Validity Study of the Primary Measures of Music Audiation.” In it, he cites a previous study, the Plymouth Meeting Study, in which the PMMA was given to 26 first and second grade students who
had studied violin for a year. The students were given the PMMA again after a semester of study, and judges scores them on three basic music performance parameters. The correlation scores ranged from .37 in the rhythm-rhythm domain, to .78 in the composite performance-tonal score domain.

Establishing these middling to fairly strong correlations went some way to establishing the criterion validity of the PMMA. That is, it demonstrated how well the items on the PMMA correspond with the construct of musical aptitude. Woodruff (1983) wanted to expand the available criterion evidence by giving the PMMA to kindergarten students before they began studying the violin, which would help to control for confounding third variables in the case of students who had already begun studying an instrument in a rigorous program.

Woodruff administered the test to 23 kindergarten-age students in Manhattan, then re-administered it twelve weeks later, controlling for the repertoire the students would learn in the meantime and which group was taught by which teacher. His predictive validity coefficients ranged from .41 to .67. This experiment featured more controls, which might help explain the more conservative range of values he obtained for predictive validity.

Gordon (1986) would go on to perform a factor analysis of the PMMA alongside the Intermediate Measures of Music Audiation (IMMA) and his Musical Aptitude Profile (MAP) based on scores obtained from 110 fourth grade students from Pennsylvania. Based on Spearman-Brown corrected correlation coefficients, the factors that were being loaded onto were determined to be stable musical aptitude, which the MAP is designed to test for in older children, and developing musical aptitude, which the PMMA and IMMA are
designed to address. The results of the factor loading of the PMMA were, as the author states, “unambiguous”. Both the tonal and rhythm scores of the PMMA loaded onto the developing musical aptitude factor while the MAP scores clearly loaded onto a different factor. (Gordon, 1986)

The factor analysis method Gordon used establishes convergent evidence for the construct validity of the PMMI, which is the degree to which a test measures the attribute it is supposed to measure, as opposed to some other underlying attribute. On the other hand, the manual also lists correlation values between the PMMA and the Metropolitan Readiness Tests, the Stanford Achievement Test, and the Lorge-Thorndike Verbal and Nonverbal Intelligence Tests. Wehner's review cites these correlations as low, and he quotes the PMMA manual as saying “It has been found that valid music aptitude tests have no more than approximately 15 percent, and usually less, of their variance in common with academic tests.” These correlations provide discriminant evidence for construct validity. That is, they show that the test is not measuring something that it should not be measuring; namely, achievement and perceptual intelligence.

Interestingly, because the test does not require any formal language skills in its execution, studies have been published examining its reliability and validity regarding students in other countries and people with special needs. Unique challenges are presented in administering the test to students of other cultures, as their music may fundamentally differ from western music in tonal structures, rhythmic structures, or both. (Stamou et al., 2009) Without going into great detail (though rigorous statistics and validity measures were applied in all cases), the PMMA was determined to be a fair measure of musical aptitude for Greek children (Stamou et al., 2009), while it was deemed fair to inadequate
for Chinese (Ji, 2012) and Korean (Lee, 2010) children. The test was also deemed insufficient in parsing out variables to determine the musical aptitude of institutionalized adults. (Gibbons, 1983)

The accumulated evidence suggests that this test is effective in determining an American child’s musical aptitude. The split-test reliability coefficients are all acceptably above .70, and, with the exception of the rhythm score, the test-retest reliability coefficients are all around or above .70. Even with a minimum test-retest correlation of .60, the rhythm score seems to demonstrate that some non-negligible measurement is being taken here.

However, no KR$_{20}$ or Cronbach’s Alpha figures could be found (outside of the manual). One of these coefficients would have been compelling evidence toward the reliability of the PMMA, because it could account for individual differences in the musical samples being presented (tempo, phrasing, dynamics, etc.) or any other intra-test variance. It seems especially unusual that one of these reliability coefficients is not calculated because the test items are in a right-wrong format; any individual item is either correct or incorrect. This fact would suggest easy calculations of KR$_{20}$ or Cronbach’s Alpha.

The low scores of every reliability statistic presented stem from the rhythm score of the test. The rhythm scores are so low on the international children’s tests that the administrators/analysts deem that portion unsuitable for children in their country. This suggests that either the rhythm portion of the test is in need of revision, or that there is some third variable in a child’s apprehension of rhythm differences that is convoluting the results of the test.

The evidence of validity is fleshed out in great detail, both in the manual and in the subsequent reviews and experiments. Gordon invented the term “audiation,” and spent
decades researching musical achievement and aptitude in children. This put him in an authoritative position to create a measure with strong content validity. He then provides compelling construct evidence in the form of a factor analysis across three different achievement/aptitude tests. As discussed above, subsequent research has yielded fair to strong correlational scores, suggesting that, while the test might not be perfect, it is a step in the right direction in examining the musical aptitude of children.

One could also argue, though, that the results of other people’s research display markedly lower scores. This might suggest some kind of biasing in Gordon’s data analyses. The test is for-profit, so one must be skeptical of claims made by Gordon or the publishing company.

Overall, though, the data is compelling. Gordon’s work stands out in a growing field of empirical research aimed at identifying children’s musical aptitudes and exploring the relationships between those aptitudes and other domains of achievement and intelligence.
References


Ji, Yue. (2012). *Usability of the Primary Measures of Music Audiation (PMMA) and the Intermediate Measures of Music Audiation (IMMA) with Elementary School Students in China*. MI: Proquest, LLC.

Lee, Jooyoung. (2010). *Usability of the Primary Measures of Music Audiation (PMMA) with 5-year-old Korean Children*. MI: Proquest, LLC.


