

The Gordon Tests

Application and Research

by Vera Mazzotta

english translation by Brent Waterhouse

Musical aptitude according to Edwin Gordon

Musical aptitude is the learning potential expressed in individuals in different ways and to different degrees, a multidimensional component, innate but determined by various environmental factors and founded on audiation⁴⁷. The accuracy and the speed with which the process of audiation occurs is the basis of each person's specific aptitude⁴⁸. Its distribution is normal⁴⁹ and the levels seen at birth, in turn determined by various elements, can be maintained or decrease depending on whether or not a child is exposed to musical experiences within 9-10 years of age. Later, with the process of myelination is completed⁵⁰, the aptitude stabilises. Non-musical individuals do not exist. Each person has a level of their own, and can express themselves musically with that level. Measuring this potential allows informed decisions to be made in guiding individuals towards musical activities, revealing their strong and weak points and optimising music education with strategies intended to develop their possibilities to the utmost. The results obtained must never be used to exclude anyone from musical practices.

⁴⁷ The ability to understand sounds even when they are not physically present (recalling them), not present for an extended period of time (listening), that are not nor ever have been present (creativity/improvisation), or to mentally organise that which one has heard in meaningful sequences and anticipate that which will follow.

⁴⁸ Even though Gordon did not believe in a correlation between aptitude and intelligence, a few studies have demonstrated that intelligence and potential are connected to each other. Rose, *On becoming and being a musician*. PhD London University 2016

⁴⁹ Gordon E.E., *The nature, description, measurement, and evaluation of music aptitude*. Chicago, GIA 1987. Normal or Gaussian distribution is a continuous probability distribution often used as an initial approximation to describe real-valued random variables that tend to converge around a single average measure.

⁵⁰ Yakolev-Lecours, *The myelogenetic Cycles of regional maturation of the brain*. In Alexandre Minkowski. *Regional Development of the brain in early life*, Philadelphia. F.Davis Co 1967

While designing the tests, Gordon bore in mind the distinction between a developing aptitude and a stabilised one⁵¹. Children whose potential is developing concentrate on a single dimension of music at a time: decisions as to tonal patterns are taken more rapidly if the latter are not organised in rhythmic structures but use notes having the same duration, and decisions as to rhythmic patterns are taken more readily if they are not tonally organised but use notes having the same pitch. A stabilised aptitude, instead, allows each person, according to their own level, to decide on one or more aspects of the pattern even when they are present at the same time.

The Tests

The Musical Aptitude Profile (MAP) was developed in 1965 for students from 9 to 19 years of age as a measure of stabilised aptitude⁵². After intense research on classifying tonal and rhythmic patterns⁵³ in order of their audiation difficulty, the following publications appeared: Primary Measures of Music Audiation (PMMA), on developing aptitude in children from 5 to 9 years old⁵⁴; Intermediate Measures of Music Audiation (IMMA), an advanced version of the previous one for children from 6 to 12 years old⁵⁵; Advanced Measures of Music Audiation (AMMA), from 12 years of age through to university; Audie for children from 3 to 4 years old⁵⁶. Other specific texts

were then prepared: Harmonic Improvisation Readiness Record, to establish whether students were able to improvise fluently, Rhythm Improvisation Readiness Record to measure their ability to adapt their improvisation to a tempo, and Instrument Timbre Preference Test to orient their choice of instrument.

The following overview solely concerns the aptitude tests. Taking into account various uncertainties and difficulties, and fully aware of its own limits, it maintains a high level of curiosity all the same, convinced of the need to experiment.

AUDIE. Audie is a test⁵⁷ specifically designed for 3-4 year-old children and is administered directly by a parent. It consists of two sub-test games, Tonal and Rhythm, each 10 minutes long and to be carried out during a single session or on different days. Each contains 10 questions, which is the highest number a child can answer without becoming distracted. One 'special song' is recurrent, and the child answers YES when they hear it and NO when they hear a different one. During standardisation, Gordon noted that the same/different paradigm used in the other tests was inadequate for very small children because, listening to patterns sung by the same voice they tended to only pay attention to it and always answer 'same'. Unlike the other two developing aptitude tests, the tonal pat-

⁵¹ Gordon, *The manifestation of developmental Music Aptitude in the audiation of Same and Different as Sound in Music*, Chicago, GIA 1981

⁵² 4th-12th grade in American education. Gordon *Manual Musical Aptitude Profile*. Chicago GIA.

⁵³ 1114 tonal and 486 rhythmic patterns. Gordon *PMMA Manual*, 1986.

⁵⁴ K-6-3rd grade: 4-9 years of age.

⁵⁵ 1st - 6th grade: from 6-7 to 11-12 years of age.

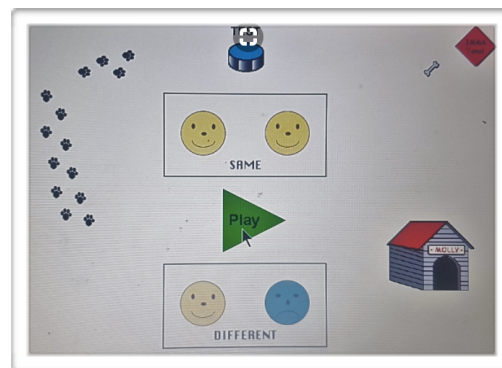
⁵⁶ Gordon. *Two new tests of music aptitude: Amma e Audie*. Newsletter Gordon Institute for Music Learning Vol. 3 n.1, 1990 and Gordon. *Advanced Measures of Music Audiation*. Gia. Chicago 1989

⁵⁷ Gordon E.E., *Introduction to Research and the Psychology of Music*, Chicago, Gia 1998

terns are heard in a rhythmic context and the rhythmic ones in a tonal context. The patterns are different either because a note is changed while the rhythm is maintained, or because a duration is modified while the melody remains the same. Children scoring 9-10 correct answers are considered to have a high aptitude, while 6-8 is medium and 0-5 low⁵⁸.

PMMA-IMMA. These two tests are similar. IMMA is an advanced version of the former, with more difficult patterns. Both contain an explanatory and preparatory/practical part. They consist of two sub-tests, Tonal and Rhythm⁵⁹, with 40 questions each in which isolated notes and durations are syntactically assimilated into tonal and rhythmic patterns. Even though the sound is synthesised⁶⁰, these patterns are presented with a certain respect for relations based on tension; the tonal ones have no rhythmic accentuation, however, and the rhythmic ones have no variations in pitch. They do not require any ability in reading: the children are asked to recognise whether the patterns are the 'same' or 'different'⁶¹ and choose a pair of images on the page/monitor.

The differences lie once again in the second pattern, since the first establishes the context. Unlike IMMA, PMMA is a test that Gordon⁶² defined as being based on immediate



impressions and intuitive responses. It presents tonal patterns made up of 2-5 notes in a major tonality, in the same keyality (C)⁶³ and at the same tempo. One of the two contains the tonic. IMMA also presents patterns in the harmonic minor, all in the same keyality, but often neither of the two contains the tonic. As a result of their own ability to internally hear the resting-tone⁶⁴, listeners infer the syntax for each tonal pattern mainly in terms of keyality and occasionally, but not necessarily, in terms of tonality. The former is an index of a developing aptitude, and the latter of a stabilised aptitude. In both tests, Rhythm includes patterns in both usual and unusual metres. A different timbre emphasises the macro-beats, which is particularly important when the patterns are different but include the same number of macro-beats. IMMA also presents many patterns in the same metre that increase the difficulty of making the comparison.



Audio 1



Audio 2

⁵⁸ G. Comeau. *Piano Pedagogy: a research and information guide*. Routledge.

⁵⁹ Gordon maintains that children in this phase can only concentrate on these two dimensions of audiation.

⁶⁰ Gordon E.E., *PMMA Manual*. Children with a developing aptitude are distracted by stimuli involving instruments that are familiar to them, but not by electronic sounds.

⁶¹ Nonetheless, in 1998 considered Gordon the possibility of using "same vs. not same", which seemed to raise the test's reliability thanks on the hypothesis of a link in small children between musical aptitude and verbal intelligence.

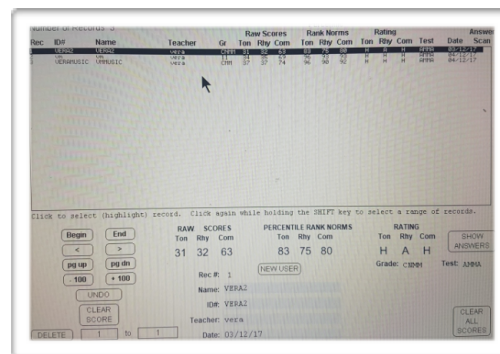
⁶² Gordon. *PMMA Manual*

⁶³ *Keyality* is the note that acts as the centre for each tonality (the tonic, in the traditional system). For example, for Gordon that which is known as B flat major is a major tonality and the keyality of B flat.

⁶⁴ The resting-tone is the syllable associated with each Tonality. C for major, D for Doric, E for Phrygian, etc.

In this case as well, the student, responding to the macro-beats they hear, infers the syntax primarily in terms of tempo and occasionally but not necessarily in terms of metre. Decisions made based on a sense of metre are a sign of a stabilised aptitude. Once listened to, a pattern may or may not be reinforced by audiation, but the pause between the two listenings, coming to 4 seconds, does not allow for the first pattern to be compared through memorisation but only recalled through audiation. The system assigns one Tonal score, one Rhythm score and one overall score that adds the two components together, coming to 40 points. Three different levels of aptitude are each represented by a range of scores⁶⁵. The scores are then elaborated as percentages according to a table that evaluates the test in relation to the level⁶⁶. The patterns are thus always the same, but are evaluated differently according to the level. It is fairly common for scores to remain within the average in both sub-tests. Contrarily, it is rare for scores to be very high in one and very low in the other. IMMA also offers a scale of overall scores, to identify children with an exceptionally high overall music aptitude, and is more precise for students with a high/medium-high developing aptitude; PMMA instead is more precise for those with an average of low aptitude. Gordon advises administering first the Tonal and then the Rhythm test, during different days of the same week if possible, and in any case no more than two weeks apart.

MAP. This test was designed, like AMMA, with the aim of more easily identifying students with a high aptitude, to



be oriented towards receiving a music education, and, for those who already benefited from one, of having a basis for predictions concerning competencies in music performance, or again, in the case of a low level, modifying the educational programmes so as to make use of the strong points and contain the weak ones⁶⁷. Given that this is a test of stabilised aptitude, it is not limited to solely tonal and rhythmic factors but takes into account the seven aspects Gordon held to be important and reliably measurable. One therefore hears melodies composed by Gordon himself in which tonal and rhythmic elements are combined. The test is divided into three sections: Tonal-Imagery, Rhythm-Imagery, Musical Sensitivity⁶⁸. Tonal and Rhythm in turn have two sub-tests, Melody/Harmony and Tempo/Metre: even though the two factors are combined, students are required to make decisions that only concern one of the two aspects. Musical Sensitivity includes three sub-tests (Phrasing-Balance-Style). In this part, acoustic instruments played by professionals and winners of important competitions, are used: the violin in Melody, Tempo and Metre and the cello for the lower voice of Harmony. In the Tonal tests, students are asked to listen to two musi-

⁶⁵ Low 20 or under, average between 21 and 79, high over 80.

⁶⁶ To gain the maximum score-percentage (99) in IMMA, for levels 4-6, 40 correct answers out of 40 are required; for levels 2-3, 38 are sufficient, and for level 1,37.

⁶⁷ Also in V. Maggini. *Verso la Music Learning Theory. Dal Musical Aptitude Profile alla formulazione dell'attitudine musicale*. Audiation Institute n.3/ 2016 pp. 12-21.

⁶⁸ Imagery: when the MAP was standardised the term audiation had not yet been coined.

cal phrases and decide whether the second is a variation of the first⁶⁹.



Audio 3

In the Rhythm tests, they are to listen to two musical phrases, and to decide whether the second is the same or different. If it is different, this is because in the second either the tempo



Audio 4

or the meter has been changed.



Audio 5

In any case, the pause between the two phrases and at times the length of the items means that their answer cannot depend on short-term memory but only on the audiation of the tempo or the metre of the original phrase. In both sections the original melody must be inferred, while at the same time a comparison arises based on the students' ability to isolate and perceive specific musical features. A huge variety of keyalities is present, for example multi-tonal or atonal using chromatic, whole-tone or octatonic scales, as well as usual, unusual and combined metres. The Musical Sensitivity section, instead, involves personal musical preferences: this is the weakest part of the test, because it is determined by cultural factors in addition to the students' specifically musical upbringing, however this may have been acquired. In the Phrasing sub-test both the violin and the cello are used. A melody is played twice, the second time changing its dynamics, phrasing, tempo or more than one of the above. Students are to decide which one sounds better. The Balance sub-test uses two melodies having an identical beginning but with different endings owing to tonal or rhythmic changes: here, students are to decide which ending completes

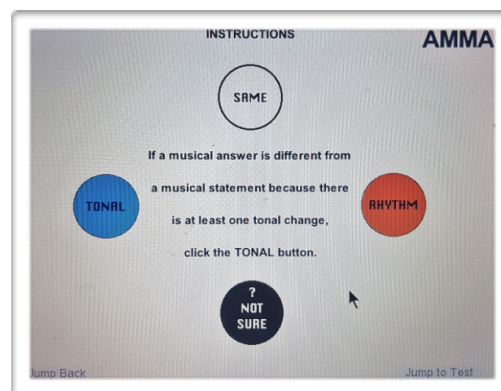
the beginning better. In the Style sub-test the same melody is presented twice but in different tempos: students are to decide which of the two speeds is more correct for this melody..



Audio 6

The first two parts are made up of 40 questions each, and the third 30. Fifty minutes are necessary for each part but the administration may take place at different moments, days or months, and even only once during schooling.

AMMA. This test was designed for students from the upper grades of Middle School through to University, in particular those who already followed music education programmes⁷⁰.



The test is similar to the MAP. It consists of 30 questions that require 20 minutes. The section involving preferences is not present. In this case as well, the examples were expressly composed, are played by professionals and do not make distinctions between rhythmic and tonal: these two aspects are heard simultaneously. Students are asked to decide whether there are differences created by a tonal or rhythmic change. While the tonal changes are caused by modifications in one or more notes, in the Tonality or in

⁶⁹ The word variation is however not used in the test; since it refers to a musical form, it is part of the field of music achievement.

⁷⁰ Gordon E.E., *Predictive Validity Study of Amma*, Chicago, Gia 1989.

the Keyality, the rhythmic changes are due to modifications in the tempo, the metre or in one or more durations. Different variants may be present but tonal and rhythmic changes never occur together. The first phrase always has exactly the same number of notes as the second, but the students cannot count or compare the number of notes: as with the other tests, the lapse of time between the two phrases is sufficient only for audiation. Even though this is a single test, scoring is divided as in the IMMA test.



Audio 7

The birth of an interest

During my internship to qualify as a junior high school piano teacher, I used these tests as one tool with which to verify the development of perceptive-creative abilities and at the same time to cultivate the ability to *internally hear music*, including vocal and instrumental performance, imitating and re-elaborating parts of the pieces played. This involved evaluating changes and improvements in the comprehension of the pieces proposed and in instrumental performance, in a group of children 10-11 years old at the end of a program in piano integrated with ear training, in turn sustained by that foreseen for formal instruction⁷¹. As a reference point, I had taken exercises from different levels of Aural Training in Practice, and Musicianship in Practice, of the English ABRSM⁷² that pro-

pose activities such as distinguishing, singing in echo, tapping, improvising, extemporaneous composition using the voice or an instrument and coordination and rhythmic recognition. Aware as I was that distinguishing, imitation and memorisation are part of audiation but taken individually are “not” audiation, I had flanked these activities with those foreseen by formal education. In order to evaluate this path of education, I used listening templates that took into account specific aspects of music education; one software in particular, Adobe Audition, to analyse the quality of the sound and timbre produced by each student; and Gordon’s tests. Since AMMA was overly complex, I chose the highest level of IMMA. After one year of work it was possible to compare the results of the tests with those of other evaluation systems: a “slight” rise in the percentiles of the tests was thus accompanied by an improvement in the other parameters considered. Given that this was a developing aptitude, in which innate potential and the musical environment continually interact and give way to fluctuations, a variation in the percentiles was probably predictable, perhaps even regardless of the type of training devised. In any case, no drops were seen. While PMMA scores can be quite unstable⁷³ given that aptitude development is affected by education⁷⁴, above all at its lower levels⁷⁵, for IMMA, above all in children undergoing the phase of stabilisation, things are more complex: in this case, it can act as a measure of a stabilised aptitude.

⁷¹ V. Mazzotta, *Valutare i risultati dell'apprendimento musicale* in http://www.musicheria.net/rubriche/?t=Valutare_i_risultati_dell_apprendimento_strumentale&p=557&f=464.

⁷² Associated Board of the Royal School of Music. <https://www.abrsm.org/it/regions/central-europe/italia/pagina-iniziale/>

The ABRSM certification is recognised in many countries and deals with the specific training of students who intend to acquire a good level of knowledge of musical language as well, by studying an instrument. Preparing the exam involves various factors, technical as well as aesthetical/musical.

The exams represent phases that allow progress to be measured based on a recognised standard of ability.

⁷³ Forsythe, *The development an Implementation of a computerized pre-school measure of music audiation*. Ph.D.diss Case Western R.University, 1984

⁷⁴ Jessup, *The comparative Effects of indirect and direct music teaching upon the developmental music aptitude and music achievement of early primary grade students*. Ph.D. diss Temple University 1984

⁷⁵ Bell, *An investigation of the validity of PMMA for use with learning disabled children* DMA diss.Temple University 1981.

This means that scores based on the number of correct answers can continue to rise as time passes, and as the children's maturity increases, while the relative scores remain constant⁷⁶. The prior data confirmed a test aimed at measuring aptitude and not achievement⁷⁷: no kind of musical training modifies the potential that distinguishes each of us, but it can contribute to bringing it to its utmost degree⁷⁸. This is why, years later and above and beyond my continuing perplexity as to the use of tests in general, I remain convinced that the lack of downward movement together with the rise in percentiles is the product of the creation of environments and educational paths based on activities that call into play the various types of audiation. The initial scores show weaknesses that somehow seemed to be related to similar difficulties in playing (use of phrasing, synchrony, fluidity, sense of rhythm). At the end of the project the variations in percentiles were however reflected by improvements in expressiveness and sound quality. This latter parameter, one of my reference points at the time, is technically tied to the different ways of lowering the key, and is thus to be considered among technical abilities. Such abilities are part of music achievement and are thus not to be confused with aptitude: the level of instrumental ability reached is not necessarily linked to one's own learning potential. Some people have a high potential but lack the technical-instrumental abilities with which to express it, and others having medium-low potential reach good technical abilities. In any case, the instrument is a carrier of one's own aptitude, whatever it may be. To explain how an improvement in percentiles within a single potential of

the children also brought about a different technical approach to music and to the instrument, I'd like to recall a phrase that my Russian teacher, with his unique way of speaking Italian, used to say when faced by technically challenging passages: *Tu pensa di musica* [You think of music]. If we think musically, with the correct élan, phrasing and agogics, our hand prepares the technical gesture which is most appropriate to express that thought. Musical aptitude is at the basis of a deep understanding of music and will thus open the way to expressiveness, and also to a search, whether personal or guided, for the technical means most adequate for expressing it.

Considerations in light of the applications

This year I resumed administering the tests in order to study their results over a longer period of time, the three years of junior high school, with children in their first year who will follow the normal path of instrumental lessons. My objective is to gain greater familiarity with the tests, to gather data, to compare it with other data and to subject it to reflection and critique. A few expedients must however be introduced. The tests must be administered individually or in a class at least twice a year, attempting each time to recreate the same conditions of silence and relaxation. The classroom should not be in a highly trafficked area (close to the school's entrance or its offices) because this would create too much noise, nor should it have sources of visual distraction. Administering the tests to the entire class makes the most of time and puts all the children in the same condition, given that the place

⁷⁶ Francisco-Soya, *Primary and Intermediate Measures of Music Audiation* in <https://www.slideshare.net/guest3b21d84d/standardized-test-analysis-project>.

⁷⁷ For Gordon, "Music achievement is a measure of what a student has already learned in music", and must not be confused with aptitude; while it is true that a high level of achievement corresponds to a high level of audiation, the opposite is not true, since students exist with high potential but low or inexistent achievement: "For example, people are not born knowing how to compose music in a given style. That must be learned. Once it has been learned, it is considered to be music achievement", *Music Aptitude and related test*, Chicago, Gia.

⁷⁸ Rises in percentile within each student's potential.

and time of the tests are constant. Rather than doing the tests on paper, I prefer to use personal computers, but if the timetable does not allow for collective administration then I propose testing two children at a time. In any case, it is a good idea to use headphones, which block out more noise. A few children, however, do not appreciate the kind of sound that comes from earphones, so it is often necessary to amplify the personal computer with a small external speaker. Whatever the conditions may be, one must maintain them and reproduce them each time. The same goes for the choice of day and time. It is preferable to get the children involved in these activities during the morning and, if this is not possible, to choose a day in which they have a lesser load of lessons. In spite of all this, many non-environmental factors that may have an influence remain, leading to a good number of variables: their concentration level, rush to finish, tiredness and anxiety. The tests presume that the time required to carry them out is always and in any case feasible for all. Today, this goes for 65-70% of the students. The so-called digital native generation has a short attention span and needs a higher amount of pauses between one task and the next than is possible during the tests. Furthermore, even though the questions are set within a sort of game⁷⁹, it is not graphically appealing or involving. Individual character inclinations, moods and environmental situations are the factors that most often intervene in perception, invalidating the results and the scores. The rhythmic sub-test is the most complex one to carry out among all the tests, and often this difficulty leads to a lapse in the students' attention and motivation. This element, which I noted for IMMA, has recently been studied for the rhythmic part of PMMA. It would seem, in fact, that wherever this test is used, outside the United States and in countries in which it is con-

sidered reliable, the rhythmic scores are always quite low. This suggests a need to revise the section in question, or that a variable in rhythmic learning may exist that somehow invalidates its results⁸⁰. MAP, with its 50 minutes for each sub-test, is a task that schoolchildren have great difficulty carrying out. For this reason, AMMA is preferable, being lighter and better adapted to a school's organisation.

Choosing the right test

The first real difficulty encountered by the teacher involves choosing the test to use. For 5-10 year-old children PMMA is available with various levels of difficulty, but if most of the students reach above-average scores then one must pass to IMMA. At the same time, for children between 9 and 12 years of age, in addition to IMMA which is partially identical to PMMA, there is MAP, which Gordon held to be more trustworthy because it was designed to provide more precise indications. Comparing the data coming from MAP, of which only the first two sections were used (Tonal-Imagery and Rhythm-Imagery) and from IMMA, one notices that they have similar elements. IMMA therefore has a certain reliability, even in cases of stabilised aptitude. PMMA, on the other hand, more closely resembles IMMA than MAP. For adolescents 12 to 13 years old or older, AMMA and MAP can be used independently of one another.

⁷⁹ The child, answering whether the two are the same or different, accompanies the dog Molly towards its doghouse.

⁸⁰ Louden. *Psychometric Evaluation of the PMMA 2013*

The reference culture – Observations

These practical difficulties must be accompanied by a few personal and general considerations. Firstly, the creation, standardisation and use of tests, not restricted to music, is the fruit of psychological studies dating to the '60s and the '70s, when educationalists, psychologists and doctors were all highly interested in defining students' initial conditions (potentiality, ability and forms of intelligence) and measuring their final results in order to determine whether the educational tools used had been effective. Secondly, standardising a test is an operation based on more or less verifiable theoretical premises; in Gordon's case, the presupposition is that musical aptitude has an average distribution. This is followed by adjustments, samples drawn from a population, the choice of a normative sample⁸¹ considered to represent it, along with templates and transformations of the scores. In this sense, one of the criticisms that has been levelled at PMMA, which I share, concerns the small reference population, made up of only 873 students from mainstream New York⁸². Reference norms obtained from a certain type of population may have no meaning, or the latter may be debatable when applied to different populations. Are we sure that these tests, as Gordon devised them, can be ad-

ministered outside the United States, in Italy as well, without changes, and that it is therefore sufficient to translate them? The PMMA for example has proven to be adequate and functional for English-speaking and for Greek children, even though in the latter case some perplexity remains as to its true reliability owing to the large differences between the scores in the study and those obtained by Gordon⁸³. It has also proven to be ill-suited to Korean and Chinese children⁸⁴. The MAP, instead, has shown to be reliable for socio-culturally underprivileged African-American students⁸⁵, German students⁸⁶, students with handicaps⁸⁷ and intellectually gifted students⁸⁸. Still today, the MAP is the reference test in the neurosciences; and yet, its use outside of the population in which it was created leaves one perplexed as to its actual applicability and the trustworthiness of its results.

Self-referentiality and interpretation

The tests are calibrated and formulated according to the same data that is later to be measured. This procedure is no doubt the result of applying mathematical formulas and statistical calculations, but it does contain an air, however slight, of self-referentiality. One error that looms

⁸¹ The group of people whose answer to the test were taken as a reference for evaluating the answers given by any other group that later took the test.

⁸² Lehman. *Review of the test PMMA*. In <http://ovidsp.tx.ovid.com/>

⁸³ Holahan-Thompson, *An investigating of the suitability of PMMA for use in England* in "Psychology of Music" 9/2 (1981) 63-68 e Stamou et alii, *Standardization of the Gordon PMMA in Greece* in "Journal of research in music education" 58/1 (2010) 75-89.

⁸⁴ Ji Y., *Usability of PMMA and IMMA with elementary School students in China*, Proquest LLC, Ann Arbor 2012; Lee J., *Usability of PMMA with 5-year-old Korean children*, Proquest LLC, Ann Arbor 2010

⁸⁵ Gordon E., *A comparison of the performance of culturally disadvantaged students with that of culturally heterogeneous students on the musical aptitude profile* in "Psychology in the school", 4/3 (1967) 260-262.

⁸⁶ Schoenoff A.W., *An investigation of the comparability of American and German forms for the MAP*, Ph.D. diss., University of Iowa 1972.

⁸⁷ Curtis C.H., *A comparative analysis of the musical aptitude of normal children and mildly handicapped children mainstreamed into regular classrooms*, Ph.D. diss., Vanderbilt University 1981.

⁸⁸ Drennan C., *The relationship of musical aptitude, academic achievement and intelligence in merit (gifted) students of Murfreesboro City schools (Tennessee)*, Ed. diss., Tennessee State University 1984.

involves not contextualising the result by putting it into relation with the knowledge possessed by those one is faced with (character, abilities, reactivity, other competences, environment). The test provides a framework that must be interpreted, raising questions as to the results. It is therefore indispensable to know how to distinguish between a measurement understood as a given score on the test and its evaluation or interpretation. If we imagine the context of the school/class, we realise how fundamental it is for the system of relations established there, emerging with their own forms of interaction, communication, culture, values, obligations, duties, representations and behaviours, not all of which are always in line with one another. In interpreting the scores, one must also bear in mind this knot of relations. Measurement is, or attempts to come across as being, objective but for it not to remain foreign to reality it needs an evaluation that is, in any case, an interpretation and, as such, is subjective. The scores cannot and must not replace human judgement; they must sustain it and provide it with reinforcement. At the same time, evaluation is necessary to make the test's framework more sharply defined. All of these considerations enable us to be critical at the very least, and to use the tests, when possible, concurrently with other reference systems⁸⁹.

Future developments

Recent research is moving towards submitting the Gordon tests to new experimentation and standardisation in the

countries in which they are applied. It is of little use to re-propose them in another language, in a somewhat commercial move that has been attempted in Italy as well⁹⁰; what is required is a serious scientific study, a longitudinal research involving musicians, psychologists, experts in psychometrics and mathematicians. The tests should be taken as a starting point, a different focus with which to orient one's work, an imperfect tool that is however connected to learning potential and from which one can glean an extraordinarily large quantity of information. A tool among other tools, towards which one must maintain a critically constructive attitude. Therefore, I believe it is necessary to continue experimenting with this branch of Music Learning Theory, maintaining an open mind and a critical outlook, because more than any other it provides us with an additional means with which to act even more responsibly, in the hope of transforming our limited experimentation into a research protocol, as has already come about in other countries.

*Develop your musicality to the highest level possible, keep playing. That which I have created can be improved. The nicest compliment I could receive would be: Gordon was a great pioneer but we have moved ahead.*⁹¹

⁸⁹ For example, a data sheet that takes into consideration a few aspects of the performance to which a precise score is to be given. Gordon himself studied the reliability of the AMMA and its ability to predict the future performance of those to whom it had been administered by submitting the performance of a study expressly composed for various instruments and voices to a commission. Various parameters were examined: well-accentuated notes, hand coordination, pedal, balance between melody and harmony, fluidity, rhythmic accuracy, choice of tempo, musical phrasing and dynamics.

⁹⁰ Gordon E.E., *Ascolta tu. Scopri le tue potenzialità musicali con i giochi di ascolto del Prof. Gordon*, Ed. Curci, Milano 2005

⁹¹ From *Edwin E. Gordon in Italy: the future of music. Interview to E.E. Gordon*, AIGAM in collaboration with Edizioni Curci 2016.

[link to the video](#)