A Response to the Colwell and Abrahams Paper
in Volume II, Numbers 1 & 2, of The Quarterly

Edwin E. Gordon

First, I would like to express my appreciation to Manny Brand, Doree Pitkin, and all contributing authors for the countless hours and thoughtful energy expended in producing Volume II, Numbers 1 & 2, of The Quarterly. I am both humbled and honored in having a publication devoted to my work, and genuinely hope the profession of music education can benefit in some way as a result of this vast effort.

After reading through each of the articles, I found six recurring general themes that require clarification since those aspects of my work appear to be widely misunderstood, or at least misinterpreted. Following these, I will record my specific responses to the matters that are of most concern to me in the Colwell and Abrahams article.

1 - There appears to be the sentiment that although I allude to supporting research in my publications, such research is not documented. With regard to more than thirty years of research in music aptitude and measurement which I have undertaken, that certainly is not the case. A perusal of the bibliographies accompanying the papers in Volume II should set the record straight. Moreover, a familiarity with the bibliography in The Nature, Description, Measurement, and Evaluation of Music Aptitudes, Chicago, GIA, 1986, offers a more detailed account.

With regard to music learning theory and related matters, admittedly, there is less documentation. What should be of primary concern to music educators is whether the hierarchical order of the sequential levels of music learning theory and the stages of audiation that I have postulated are reasonable and logical. To take the time to compare my ideas with those of others and to declare music learning theory the winner has little interest to me. It is the process more than the product of learning that is compelling. Thus, particularly over the past ten years I have been engaged in teaching music to students of various ages, including children eighteen months old. From such experiences and from observing the work of other teachers who do and do not follow my pedagogy, I have developed insights into the validity of my theories and I have more than occasionally adapted my writings to current findings. I have never attempted to publish that empirical research because, having been a journal editor myself and having worked with editorial boards, I know the reports would be rejected on traditional grounds; they include no classical control groups, no tests of statistical significance, and so on. My research in the multi-faceted dimensions of music aptitudes, however, has offered indirect knowledge about the structure of music learning theory and the potential for the development and acquisition of audiation skills. In reality, students throughout the country who are exposed to traditional music education have served as my control groups.

Traditional comparative research to persuade critics that music learning theory is superior to more commonly used methods is not realistic, because although such experimental research may be theoretically designed in accordance with what is usually taught in university classes, the results would be less than credible. Consider the possibility of finding at least two teachers who teach similar groups of students in similar environments who are capable of teaching well both music learning theory and a method to which it is to be compared. If a significant difference were found between the groups of one teacher who has taught only music learning theory and another teacher who has taught only the other method, is it impossible to discern whether that difference is a result of one method having been better than the other or of one teacher having been better than the other. In the case where one teacher is assigned to teaching both groups, the teacher is routinely criticized for being partial to the method in which students demonstrated significantly higher achievement. Given adequate funding and the prestige that music education should have, many replications in numerous locations including many teachers over a period of years would need to be undertaken to come to what is referred to as "proof." Moreover, as with the results of doctoral dissertations reported in Volume II, non-significant differences are typically the result unless an experiment is conducted over a period of at least one academic year, and preferably two.

Throughout my career I have not been able to identify just two teachers under the desirable conditions described above to compare the efficacy of even Orff and Kodaly methodologies. Consider the dozens upon dozens of studies that would be needed to carry out the type of research called for involving a comparison of a traditional method with given sequences of a series of multiple levels of music learning theory or with instruction at the various stages of audiation. Then, too, there are the more practical problems of scheduling; securing permission from parents, students, administrators, and research committees; and convincing oneself that it is professionally responsible to have a group of students taught in a manner that he or she believes is not the most desirable. It is interesting to note that the type of comparative research endorsed by some of the authors is work designed primarily by graduate students who, as they themselves imply, engaged in impractical but theoretically correct training exercises. Rather than relying on that research and asking me to engage in unproductive research in addition creating the theory itself, would it not be of greater benefit to the profession if mature and conscientious music educators, who are no longer graduate students, would attempt to engage in the actual research themselves? I, for one, would be grateful should that occur.

A final word about the research issue. There are practical significance and statistical significance. Given the dearth of valid standardized criterion measures in music education that have established norms, researchers are forced to rely solely on statistical significance. Assuming that there were no theoretical and practical complications in designing comparative research to prevent a statistically significant difference between groups from being found, how would the problem of practical significance be addressed? It would not be a credit to music education if most, if not all, of the research results were interpreted in terms of only statistical significance, with practical significance being ignored.

2 - A student's attitude toward music is, of course, important. My thesis is simply that students are motivated through music learning theory partly because it provides a curriculum that develops an individual's music skills and engenders in students a sense of accomplishment and success. A good teacher creates good attitudes by teaching music and developing music skills in one's students and not by teaching about music or by constructing artificial situations in the hope of developing good attitudes.

There is a difference between interest and motivation. Interest is generated outside the student. It may dissipate as quickly as it was acquired. Interest does not always culminate in motivation. Motivation is generated from within and sustained by the student himself or herself, depending upon his or her attitude about what he or she is learning. Learning has more to do with students' attitudes than does teaching.

3 - A fundamental principle of learning sequence activities (those that take place during only the first ten minutes of a class period or rehearsal in a music learning theory program) is for students to learn tonal patterns separately from rhythm patterns. When a tonal pattern and a rhythm pattern are combined and presented as a melodic pattern, the student will be less able to conserve and generalize the tonal pattern to other rhythms or to generalize the rhythm pattern to other melodies. The student might actually be prevented from acquiring vocabularies of tonal patterns and rhythm patterns that are necessary to learn to audiate and to audiate to learn.

In learning sequence activities associated with music learning theory, tonal patterns and rhythm patterns are not taught in isolation of music syntax. That is, before students hear and perform tonal patterns in learning sequence activities, they hear literature performed in the tonality of the patterns. Then, immediately before the tonal patterns are presented, the teacher performs a short series of tones, called a sequence, to establish the appropriate tonality before the patterns are to be audiated and performed. The case is similar for rhythm patterns. Before students hear and perform rhythm patterns in learning sequence activities, they hear literature performed in the meter of the patterns. Then, immediately before the rhythm patterns are presented, the teacher performs a short series of durations, called a sequence, to establish the appropriate meter before the patterns are to be audiated and performed. Music syntax is not only inherent in the teaching of patterns in learning sequence activities, it is found, of course, in traditional classroom and performance activities that are presented after the first ten minutes of instruction. Learning sequence activities provide curricular structure, accountability, and precise conditions for teaching to students' individual musical differences.

4 - Learning sequence activities are criticized as being "drill." When taught pedantically, without enthusiasm, by a teacher with limited musicianship, activities reminiscent of drill become the result. When taught by a zealous, musical teacher, learning sequence activities are not even remotely associated with drill. It is well to remember that drill is a method of teaching not an outcome of or an attitude toward learning.

5 - Perhaps because of tradition, some music educators cannot come to terms with not coordinating tonal patterns and rhythm patterns taught in learning sequence activities with literature students are performing in classroom and choral and instrumental ensemble activities. My reasons for not recommending such coordination have been well documented and researched in the ways I have described. They are the following:

> When patterns are selected on the basis of the frequency with which they are found in literature, the difficulty levels of the patterns and, consequently, students' individual musical differences are ignored.

When literature is chosen on the basis of the difficulty levels of the patterns it includes, quality of literature is generally sacrificed.

When a pattern is identified in literature, it is in a melodic context. That is, it combines both tonal and rhythm elements. Thus, it does not serve to establish either a tonal pattern vocabulary or a rhythm pattern vocabulary. Given typical scheduling and administrative problems, not to mention conservation and the incomprehensible number of combined tonal and rhythm patterns that would need to be taught, the establishment of a melodic pattern vocabulary would seem almost impossible.

In music learning theory, it is basic that tonalities and meters be introduced in classroom and performance activities before patterns are heard and performed in those tonalities and meters in learning sequence activities. To wait until certain patterns are ready to be taught before allowing students to begin to hear and perform literature in a variety of tonalities and meters in which the patterns might be found is to delay the development of audiation skills. The longer that delay, the less able a student will be to develop audiation skills.

When students are continually taught to "discover" familiar tonal patterns and rhythm patterns in literature, they are being deprived of developmentally learning how to make such inferences on their own. Such generalizations, which relate to creativity and improvisation and to what is commonly referred to as sight reading, form the basis for musical understanding and the acquisition of advanced audiation skills. In music learning theory, discrimination learning provides the readiness for inference learning. That is what an understanding of music is all about.

6 - Audiation is often confused with imitation, inner hearing, aural perception, memorization, and recognition. It is possible to imitate, inner hear, aurally perceive, memorize, and recognize without audiating. Audiation is to music what thinking is to language. A simple definition of audiation might help: Audiation is the ability to hear **and to comprehend** music for which the sound is no longer or may have never been physically present. To the extent that a student can audiate, he or she will be capable of achieving in music. For a comprehensive clarification of the types and stages of audiation, may I suggest reading two of my books: <u>Learning Sequence in Music: Skill, Content, and Patterns</u>, Chapter Two, Chicago, GIA, 1998, and <u>A Music Learning Theory for Newborn and Young Children</u>, Chapter Three, Chicago, GIA, 1990.

The Colwell and Abrahams Article

Quotations from the Colwell and Abrahams paper, without attending to choice and use of words, are underlined. My responses follow immediately below each of their statements.

Page 19: Gordon is a behaviorist. He believes that learning progresses through successive stages, and he uses psychological principles to explain music learning. He has not accepted the cognitivist philosophy of learning, and he stresses the importance of each student achieving his or her full potential.

A behaviorist psychologist such as Watson, Thorndike, or Skinner would probably not agree that, based on the simple description above, I adhere to behaviorist principles. Quite frankly, and regardless of what one may believe, I consider myself to be more in line with cognitive and developmental thinking than with behaviorist principles. Don't psychologists of different persuasions use psychological principles to explain learning? The following sentence written by another author on page 94 of Volume II may in part underscore the futility of labeling persons: "It should be noted that Gordon is not a behaviorist because he recognizes the limitations of environmental influences."

Page 19: In some respects, his work is rather specialized, and a student taught by the "Gordon system" is more likely to demonstrate achievement on one of his tests.

There is no Gordon system. As explained throughout my writings, music learning theory is designed to embrace many methods that are based on sequential objectives, each objective serving as a readiness for achieving the next objective. An attractive aspect of music learning theory is that a teacher is encouraged to adapt it to his or her own personality and teaching style. And, of course, I would expect that a student who is exposed to learning sequence activities or any other musical approach to music learning could be expected to score to his or her fullest potential on one of the six levels of the Iowa Tests of Music Literacy, which is my **only** achievement test. A valid music aptitude test measures a student's **potential** to achieve in music. Only a valid music achievement test can measure a student's music achievement.

Page 20: The flaws in his work are often a result of Gordon's attempt to "teach" the reader or user of his research rather than carefully and objectively presenting and interpreting the data. Even in the manual for his stabilized music aptitude test, the Musical Aptitude Profile (MAP) (1965), Gordon devotes pages to suggestions about what should be done by the teacher for the student who scores below average on the test. If there is one student in a school whose score is less than it might be because his or her aptitude has not stabilized, Gordon wants the teacher to begin an instructional campaign to change that student's score before it is too late.

It is not uncommon for music educators to request researchers and test developers to interpret the practical implications of their research and to explain how their tests might best be used. I endeavor to meet those requests in my work in addition to carefully and objectively presenting data. I cannot believe that one might consider my test manuals not to be written carefully and objectively. As a matter of fact, the typical complaint is that my manuals contain too much data.

In the manual for the Musical Aptitude Profile, which, by the way, was revised in 1988 and has been used for more than twenty five years throughout the world, as well as the manuals for my other tests, I offer suggestions for teaching to students' individual musical differences. Adapting instruction to individual needs is a cornerstone of music learning theory. Thus, I offer suggestions not only for adapting instruction to the needs of low scoring students, but to average and high scoring students as well. Without such concern in instruction, high scoring students become bored and low scoring students become frustrated.

A student's score cannot be "less than it might be," on the Musical Aptitude Profile because it is a test of stabilized music aptitude. Whether a student is in the developmental or stabilized stage of music aptitude, he or she may obtain any score from very high to very low on an appropriate test. As explained in the test manual, a student should not be administered the Musical Aptitude Profile unless his or her music aptitude has indeed stabilized. And, the test need be administered only once to a student because his or her music aptitude has stabilized.

I do not recommend campaigns of any type. Colwell and Abrahams should understand that a student's valid stabilized music aptitude score cannot be practically changed "before it is too late." Only developmental music aptitude scores are affected by practice and training and other environmental effects. Supportive research, conducted in the classical manner, is well documented. As examples, with regard to stabilized music aptitude, see pages 40 through 43 of Edwin Gordon, A Three-Year Longitudinal Predictive Validity Study of the Musical Aptitude Profile, Studies in the Psychology of Music, Volume V, Iowa City, University of Iowa Press, 1967, and with regard to developmental music aptitude, see pages 25 and 32 through 37 of Edwin E. Gordon, The Manifestation of Developmental Music Aptitude in the Audiation of "Same" and "Different" as Sound in Music, Chicago, GIA, 1981.

Page 20: He relied heavily on correlation techniques, and a correlation study is vulnerable to several criticisms. One correlation has been of special interest to Gordon, and that is the relationship of his aptitude test scores with measures of intelligence.

A clarification of the vulnerable criticisms would be helpful. Are Colwell and Abrahams referring to the homogeneity of groups of students, to the quality of the Musical Aptitude Profile, or to the conditions under which the test was administered. Certainly any good researcher is aware that relationship does not imply causation. That is why correlation techniques are used to determine the **longitudinal** rather than the **concurrent** relationship between test scores and success in instrumental music.

The relationship between music aptitude and intelligence test scores were not of special interest to me. They were of preliminary interest. As explained in the test manual, because the correlations between music aptitude and intelligence tests were found to be low, it was only then considered advisable to embark on the more scholarly three year longitudinal predictive validity study that was so expensive in time, energy, and money.

Page 21: One must ask if Gordon's achievement test was a valid indicator.

An indicator of what? As a matter of fact, several validity criteria were used. It was found that specially constructed tests of music achievement and etude performances evaluated by independent judges correlated higher with music aptitude scores than did teachers' ratings. Colwell and Abrahams suggest that the reader should "wonder about" those facts. I hope and suspect that many readers have. That is one of the purposes of engaging in research and publishing results.

Page 21: In order to document the power of the MAP, Gordon selected the upper and lower ten percent of the students (as measured by his test) and compared the mean scores of the two groups on the various achievement tests and teacher ratings. He obtained a difference between these extreme groups but he also should have conducted a similar study using the top and bottom ten percent of the students as identified by teacher ratings.

Upon reading the quotation again, it will be apparent that Colwell and Abrahams say I did what they say I should have done with regard to teacher ratings. I call attention to the data presented on pages 19, 27, and 37 of my monograph, A Three-Year Longitudinal Predictive Validity Study of the Musical Aptitude Profile, Studies in the Psychology of Music, Volume V, Iowa City, University of Iowa Press, 1967. The data are there.

Page 21: As Gordon is a pioneer in careful music research, one is not prone to disregard his conclusions. Most music educators, however, think that a competent teacher working with the same children constantly on instrumental instruction over a period of three years would be better able to judge musicality and music achievement in a child than an impartial judge making a judgment based on two hearings of three etudes. The possibility exists that Gordon's objectives for the aptitude test and for the achievement test were similar.

The point of that specific aspect of the research was to indicate to music teachers that an objective analysis of their students' music aptitude could serve as an aid to their subjective judgment about their students' music aptitude. Because most music teachers assess aptitude on the basis of achievement, it is no wonder that the correlations were found to be low. Is the purpose of research solely to confirm old beliefs? If so, why do it? Is there not room for research to make new discoveries and to challenge old beliefs? I should explain that two judges, not one, listened to the etudes three, not two, times. With regard to the final sentence, perhaps it was intended for another paragraph.

Page 21: The correlations between composite test scores and teachers' estimates ranged from .64 to .97, an impressively high correlation. . . these correlations are sufficiently high to provide assurance to instrumental teachers of their professionalism.

The fact is that the correlations between composite test scores and teachers' ratings range from .35 to .39, far below the correlations for other achievement criteria. I call attention to the data presented on pages 18, 25, and 35 of my monograph, A Three-Year Longitudinal Predictive Validity Study of the Musical Aptitude Profile, Studies in the Psychology of Music, Volume V, Iowa City, University of Iowa Press, 1967. The low correlations between teachers' estimates of their students' music aptitudes and the students' Musical Aptitude Profile scores have also been demonstrated in other studies. For example, see my research report, "Taking into Account Musical Aptitude Differences Among Beginning Instrumental Music Students," American Educational Research Journal, 18, 1970, 195-213. It baffles me that data might be construed in such a way as to suggest to an instrumental music teacher that he or she is unprofessional. The point of that research was not to degrade teachers. The point was to apprise teachers of the fact that because a valid music aptitude test hears what a teacher cannot see, they might want to investigate the possibility of using such a test for the improvement of instruction.

Page 23: Instrumental instruction did not affect scores on MAP. . . The finding of lack of impact on instruction is contradicted in Gordon's later writing about developmental aptitude and stabilized aptitude.

I have maintained in my writing that stabilized music aptitude is not affected by practice and training. I have also maintained in my writing that developmental music aptitude may be affected by environmental factors, positive and negative. That is not a contradiction. It is an explanation of how two types of music aptitude are different from each other.

Page 23: Gordon believes that he could not compare the effectiveness of two instructional methods without knowing the student's aptitude because a researcher might mistakenly reject a method; the apparent failure of the method might be due to lack of talent in the subjects.

What is alleged is not what I believe. If a teacher is unaware of students' music aptitude scores, the possibility of rejecting a method in an experiment would apply equally to all methods. That the failure of a method might be due to a lack of "talent" (I presume what is meant is aptitude) is certainly remote. All students have at least some music aptitude. And, if the subjects were randomly assigned to groups, as they should be, no such fear is warranted.

Page 23: The reliability indicates that students can successfully answer most questions, as all of the sensitivity items have a difficulty index of .70 and above.

The reason a test developer investigates item characteristics is to learn something that reliability coefficients do not specifically reveal. A test can be reliable whether it is easy or difficult.

Page 23: The discrimination indicies are higher than would be expected with such easy questions. Either there is an error in the manual or the test could be shortened considerably without loss of validity.

If 50% of high scoring students answer an item correctly and 50% of low scoring students answer that item incorrectly, the test item would have perfect discrimination. The test item would have a difficulty level of 50%. Given a test item of 50% difficulty with perfect discrimination, a test item of 70% difficulty could have at least a high discrimination value. I doubt that there is "an error" in the manual.

Moreover, if it were found, for example, that all of the discriminating items in a test have "different" as the correct answer and all of the non-discriminating items have "same" as the correct answer, it would be folly to discard all of the non-discriminating items and believe that validity would not be sacrificed. In professional test development, there are content, construct, and process validities to be considered in addition to the characteristics of the items. Item indexes provide information about preliminary validity, not validity itself.

Page 23: Gordon has not continued his interest in stabilized aptitude since the data on MAP were gathered. His interest is now focused on developmental aptitude, a term that can be interpreted as early achievement.

That is simply not true. My stabilized music aptitude test, Advanced Measures of Music Audiation, was published in 1989 by GIA, Chicago. It was designed for use primarily for college and university students. Further, just last year, in 1990, my monograph, A One-Year Predictive Validity Study of the Advanced Measures of Music Audiation, was published by GIA, Chicago. Colwell, if my memory serves me well, actually assisted in the standardization of the test.

If one insists, developmental music aptitude may be interpreted as early music achievement. It must be remembered, however, that the level of music achievement that a student is able to attain is determined by his innate potential. Good instruction lessens the effects of biological limitations.

Page 24: Whether the selection of easy or difficult patterns makes any difference in measuring music aptitude is unknown; the length of the pattern is presumed important.

As explained in the test manual for the Primary Measures of Music Audiation and the Intermediate Measures of Music Audiation, what differentiates the two tests is not found in their design, only in the difficulty levels of the patterns that are included in each test. Only easy patterns are used in the Primary Measures of Music Audiation. A comparison of the notated patterns and their relative difficulty levels presented in appropriate tables in the test manual will quickly assure one that the length of a pattern is indeed **not** important to its difficulty level, presumed or otherwise.

Page 24: He accepts at face value that tone and rhythm are the two major components of aptitude; however, in his tonal and rhythm pattern research, the "ability of fourth grade students to hear pairs of patterns as being the same or different was found to have virtually no correlation with stabilized music aptitude as measured by the Musical Aptitude Profile" (1981, p. 73.) This statement could be interpreted to mean that there is almost no relationship between developmental music aptitude and Gordon's definition of true music aptitude. The same data (1979, p. 90) are used, however, to argue that scores from the PMMA are related to the MAP: . . .

I do not accept simply at face value that tonal and rhythm dimensions are the two major components of developmental music aptitude. As explained in the test manual, it was found in research in the development of the test that measures associated with timbre, loudness, and preference were too unreliable to use. As a result, only the two most reliable measures, tonal and rhythm, were used. With the use of "however" notwithstanding, what that statement has to do with the remainder of the paragraph is beyond my comprehension.

With regard to the remainder of the paragraph, it should be understood that the methodology of determining the difficulty levels of patterns was made to be purposefully different from the design of an aptitude test. Thus, it should not be surprising that the correlation between pattern difficulty levels and aptitude scores might be low. To argue that because scores on a stabilized music aptitude test are not highly correlated with results from the pattern research study, scores on a stabilized music aptitude test and a developmental music aptitude test cannot demonstrate a moderate correlation, is not to understand the nature of correlation coefficients. Actually, the moderate congruent validity correlations between the two types of music aptitude tests are ideal and suggest, not support, the idea that the goals of the two tests have commonality. Had they been much higher, it would need to be questioned whether there are in fact two types of music aptitude.

I would be obliged to Colwell and Abrahams if they might direct me to where I use the term "true music aptitude" so that I can discover how I defined it.

Page 25: Concurrent validity is dismissed because of the difficulty of distinguishing between informal and formal music achievement on the one hand and innate music aptitude on the other (1979, p. 86). Our point is that there can be no reliable criterion.

First, the reference should be 1978, not 1979.

When the following introductory sentence which precedes the adaptation of my words is not deleted, my point retains its logic. "Because developmental music aptitude, by definition, is in a constant state of variation, it would not be logical to seek criterion-related validity for the Primary Measures of Music Audiation."

Of course there is not only a reliable criterion measure, there is a valid one. It is an integral part of longitudinal predictive validity. There have been several such studies undertaken, some of which are made reference to by Colwell and Abrahams. In terms of test development, concurrent validity is only a stepping stone to longitudinal predictive validity. Given longitudinal predictive validity, concurrent validity pales in comparison.

Page 25: The correlation between the two parts of the test of about .50 was confirmed by our administration of the test, and similar figures were found in at least two other studies.

It is not clear from what precedes and what follows that statement whether Colwell and Abrahams are referring to the reliability of one of the tests or to the intercorrelation between two of the tests. Either way, however, their point is not well taken. If a test is not administered with care to achieve a desirable and unbiased high reliability, a low reliability, even lower than .50, will probably be the intended result. When two tests in a battery have much more than 25% in common, as suggested by a coefficient of .50, the content and construct validity of one or both of the tests comes into question.

Page 25: These unlikely, but possible, figures occur only when a test resembles a Guttman rating scale. . . Gordon's statement that "there is a considerable range of item difficulty levels for each test in all grades" (1979, p. 70) cannot be taken seriously; one difficult question (#19) hardly gives "considerable" range; it is the only one with a difficulty index of less than .5 for third grade students. . . These data, as suggested earlier, tend to explain the results Gordon obtained in his factor analysis as well as some of his other statistical data.

If my figures are possible only when a test resembles a Guttman rating scale, and the Intermediate Measures of Music Audiation are by far not designed as a Guttman rating scale, how can the figures be possible? I sincerely hope that in their contradictory statement Colwell and Abrahams are not questioning my integrity and suggesting that I am untruthful with data.

With a careful examination of Table 6, (not Table 5 as mistakenly reported by Colwell and Abrahams) of the 1982 edition of the test manual, it will be discovered that there is **not** only one question (#19) with a difficulty level less than .5 for third grade students. As a matter of fact, there are three in the Tonal test (and one of them is not #19) and there are two in the Rhythm test. Colwell and Abrahams read the wrong column. Moreover, with regard to a considerable range of difficulty levels, I call attention to Table 5, specifically for third grade students, questions 4, 15, 20, 28, 31, 38, and 40 in the Tonal test and to questions 7, 14, 19, 20, 27, 29, 30, 34, 35, and 37 in the Rhythm test. Once the table is read correctly, can doubt remain that the questions cover a considerable range of difficulty? Perhaps I can be taken seriously and my "factor analysis" and "other statistical data" may now be interpreted more clearly, particularly with the understanding that it is reliabilities, intercorrelations, and item discrimination values that are primarily responsible for factor analytic outcomes, not item difficulty indexes.

Although Colwell and Abrahams make no mention of it, I think it advisable to inform the reader that the PMMA and IMMA manuals were revised and combined in 1986. Manual - Primary Measures of Music Audiation and the Intermediate Measures of Music Audiation, Chicago, GIA.

Page 26: <u>The test functions as a criterion-related achievement test with a ceiling that is rather easily attained. . . Teachers with strong music programs will find the norms tables for third grade students to be disconcerting . . . The table of norms in the manual suggests that the test is inappropriate for third grade students (too easy), but these are likely underestimated . . . Accordingly, the test is not appropriate as an aptitude test for all the recommended ages.</u>

Space does not provide for an explication of a criterion-referenced test. (The term "criterion-related" is traditionally associated with test validity.) I can assure the reader that I have never written a criterion-referenced test. As a matter of fact, a criterion referenced aptitude test is an impossibility.

As explained in the test manual, the <u>Intermediate Measures of Music Audiation</u> were written because need demanded it. Many students in the developmental music aptitude stage who are exposed to learning sequence activities, as well as other methodologies that emphasize a type of audiation, are able to raise their developmental music aptitudes dramatically. As a result, it is recommended in the manual that when a substantial number of students score high on the <u>Primary Measures of Music Audiation</u>, the use of that test should be discontinued and the <u>Intermediate Measures of Music Audiation</u> should be used in its place. In some schools with excellent programs, it is even necessary to use the <u>Intermediate Measures of Music Audiation</u> beginning in the first grade.

In passing, it might be pointed out that the very fact that the need has arisen for the <u>Intermediate Measures of Music Audiation</u> itself provides objective research which supports the value of music learning theory and learning sequence activities.

Page 26: <u>Gordon is a good teacher, but he does make unsubstantiated statements in what is presented as a scholarly test manual: . . . These (pedagogical) statements are often logical, but none is supported by research.</u>

I believe that it is possible to write a test manual in a scholarly manner and at the same time to offer direction to a teacher for interpreting test results in the form of teaching suggestions. As explained heretofore, the teaching suggestions that are offered are based on my research that has been and is being accomplished in the classroom as I teach students. Perhaps I am egotistical to think that teachers might find the information I have gleaned from my experience to be of some use to them.

Page 27: <u>"It has been found that the timbre of the music instrument a student plays is second only to his music aptitude as an important factor in instrumental music" (Gordon, 1986, p.5). We know of no confirming research for this statement.</u>

The "confirming" research, designed and conducted by me, may be found in two sources, with more forthcoming. They are "Final Results of a Two-Year Longitudinal Predictive Validity Study of the Instrument Timbre Preference Test and the Musical Aptitude Profile," <u>Council for Research in Music Education</u>, 1986, 89, 8-17 and <u>A Two-Year Longitudinal Predictive Validity Study of the Instrument Timbre Preference Test and the Intermediate Measures of Music Audiation</u>, 1989, Chicago, GIA Monograph Series. Colwell, I believe, was editor of the journal in which the first study was published.

Page 29: Whenever Gordon refers to duple or triple meter, he refers not to meter signatures which account for commonly understood duple or triple groupings of pulses, but to the division of pulses. . . There is no argument among theorists that the traditional classification is inconsistent and unable to deal adequately with much music.

The first sentence is true, but I resist the use of the word "pulse." I believe that a measure (not a meter) signature is no more able to indicate meter than a key signature is able to indicate key. Just as the key signature of one sharp may indicate G keyality if the music is in major tonality, E keyality if the music is in minor tonality, A keyality if the music is in dorian tonality, and so on, a measure signature of 2/4 may be used to indicate duple meter, triple meter, combined meter, and even unusual meter. In a word, notation, including measure signatures, cannot explain. Notation can only assist a musician in recalling what he can already audiate. Because music is written with two "beats" in a measure does not necessarily make it duple, and because music is written with three "beats" in a measure does not necessarily make it triple. Measure signatures are enrhythmic just as key signatures are enharmonic.

The sections titled "Methods of Instruction" and "Classification Systems" on pages 28 through 31 in which Colwell and Abrahams devote considerable space to examining my ideas about rhythm convey misunderstanding and seem misguided. Rhythm cannot be explained through either music theory or notation. The ability to engage in artistic movement is a prerequisite, a readiness, for understanding rhythm. I would suggest that Colwell and Abrahams participate in a seminar that I offer on rhythm. Only in that way might they be able to examine objectively their own traditional beliefs and open their minds to new ideas. If they remain hostile to my ideas, at least there is the possibility of having a rational discussion which might lead all three of us to the "truth." In the interim, may I suggest that if only music theory can offer guidance to them in examining the nature of rhythm, I recommend that pages 30 through 33 be read in Étienne Loulié, Elements or Principles of Music, translated and edited by Albert Cohen, New York, Institute of Mediaeval Music, Ltd., 1965. Loulie, in the seventeenth century became interested in the matter, recognized the problems of what was even then referred to as simple and compound meters.

With regard to the last sentence of the quotation, my experience tells me that there remains considerable disagreement among music theorists about rhythm. If there were not, why are some still writing about simple and compound meters in their textbooks, even though, when challenged, they cannot demonstrate simple and compound meters in movement?

Pages 31-32: Gordon compares his book with Mursell's "The Psychology of School Music Teaching (1931), but Gordon's contribution is not as comprehensive.

In the Preface of The Psychology of Music Teaching, I make the following statement: "In 1931, James Mursell wrote in The Psychology of School Music Teaching, ''There never was a time when music education more urgently needed the help that scientific psychology can give.'' It can be safely stated that almost forty years later, music educators' professional needs are not materially any less exigent. Hopefully, this book will achieve its intended purpose in addition to that of providing impetus for designing and conducting relevant experimental investigations.''" It stretches the imagination to construe from the above that I made a comparison of my book with Mursell's and Glenn's. That was never my intent, and to the best of my knowledge, it was not the intent of Charles Leonhard, the editor of the series in which the book appeared.

Page 32: Although Gordon holds the position of Carl E. Seashore Professor of Research in Music Education at Temple University, Seashore's ideas on music aptitude receive no more than a few pejorative comments from Gordon because of Seashore's elemental approach to music aptitude. . . The authors of this article make note of this interesting comparison in that Gordon's teaching ideas are among the more "elemental" in 1991, focusing on the mastery of rather small patterns.

After re-reading pages 12 through 15 in my book, The Psychology of Music Teaching, which represent a relatively large segment devoted exclusively to the music aptitude work of Carl E. Seashore, I find no disparaging remarks about Dean Seashore or his work. To those who might have reason to infer such negative implications from my professional analysis of his aptitude tests, I extend my apologies to them as well as to the entire professional community. I have always held Dean Seashore and his work in high esteem. How foolish it would have been for me to request, as I did, that Temple University name the Chair I was to occupy in honor of Carl E. Seashore if I did not respect his pioneering work in the psychology of music. Without the work of Dean Seashore to guide me, I doubt that I would have attained even the meager amount of success that some may believe I have.

There is a difference between using patterns for testing purposes and using patterns for instructional purposes. For example, as I have explained earlier, in learning sequence activities, tonality and meter are established before students are asked to listen to and to perform patterns. That is not the case with a developmental aptitude test in which students only listen to patterns and are required themselves to establish syntax for the patterns. Moreover, students do not master patterns in learning sequence activities, they audiate them.

Page 32: Gordon would want reviewers to recognize his earlier contributions to music education, which we have attempted to do, but Gordon modifies his definitions and ideas with some regularity, and knowing his present position on music teaching and learning is difficult.

Yes, I continually learn from my research, and even though I am in print, I have the courage and obligation to change my position when research results so direct. My present position can always be located in my most recent writings, such as A Music Learning Theory for Newborn and Young Children, which Colwell and Abrahams cite.

Page 32: He maintains, however, that a student cannot progress beyond the musical limits established by his or her inherited ability.

I have never said that. I do not use the word "inherited" in that context. When a capacity is inherited, it can be predicted through genealogy. None of my research suggests that music aptitude is inherited. My research does suggest that music aptitude is innate. That is, for whatever the random reason or reasons, one is born with at least some given level of innate potential for learning music.

Page 33: Developing one's potential, then, is the damage-control program that is initiated after birth; . . .

I do not think of appropriate instruction in education as "damage-control." Negative and militaristic thinking of that type ignores the compelling and over riding issue of gathering information about how we learn when we learn music.

Pages 34 - 35: Reimer states (1989), "An argument has been made that programs should stem not from a philosophy, but from a psychological theory of how children learn or from learning theory (a term since abandoned by psychologists)" (1989, p. 149). In suggesting that learning theories cannot provide the structure for the learning process, Reimer suggests that psychology can be applied to anything from music to housecleaning, but psychology cannot differentiate between the teaching of cooking and the teaching of music.

Anyone who is current in music education is aware that Bennett Reimer and I disagree about many aspects of music education. I do not believe that psychology or philosophy alone, one without the other, can solve the myriad of problems facing music education. Nor do I believe that music learning theory is outworn and useless. My work and that of others in the classroom coupled with research findings suggest to me that music learning theory is a vital force in contemporary music education, the proclamations of philosophers and psychologists notwithstanding.

* * *

In light of the quotation of Colwell and Abrahams, The stimulation of a profession comes from wrestling with ideas, from the process of sorting out the claims made by the best minds in the profession," it is my hope that these responses will offer integrity and responsibility to that process.