Order Number 9132372

A survey of some of the theories that have developed in the field of music aptitude testing and research since 1940

Rothe, Penelope Nichols, Ed.D.

Harvard University, 1991

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A SURVEY OF SOME OF THE THEORIES THAT HAVE DEVELOPED IN THE FIELD OF MUSIC APTITUDE TESTING AND RESEARCH SINCE 1940

PENELOPE NICHOLS ROTHE

An Analytic Paper Presented to the Faculty
of the Graduate School of Education of Harvard University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

1991

TABLE OF CONTENTS

DISSERTATION ABSTRACT: A Survey of Some of the Theories That Have Developed in the Field of Music Aptitude Testing and Research Since 1940. Submitted by Penelope Nichols-Rothe 4/1/91.

The Research Question

How has research in music aptitude changed over the last 50 years, what theories have influenced it, where is it going?

Relevant Theories

Early researchers in music aptitude were influenced by Elementalist and Gestaltist views. More recently, new research inspired by cognitive and developmental theories created new interest in music perception and aptitude. This new research contains elements of the old theories but also reflects and informs new theories of how people perceive music, as well as other forms of knowledge, in psychology today.

Answers to the Question the Research Might Produce

1) Theories about music perception have come out of the general psychological theories prevalent at the time the research was done. 2) New theories have changed the way research in music perception is designed. 3) Standard methods of designing experiments may not be adequate to research the nature of music perception.

Methods and Types of Analysis Used and Data to be Gathered:

A historical survey of the theories influencing the design of music aptitude tests and descriptions of the tests themselves. Comparative analysis of test designs, methodology, reliability and validity reports, and theories underlying the design of the older tests and newer experiments. Critical analysis of pitch subtests and generation of further questions to pursue in research. Data was gathered through historical accounts, journal articles and books, published and unpublished research reports, dissertations and from my own follow-up study of one experiment.

Summary, Defense, Justification

This historical review will help to account for: directions that have developed; new ideas for further research; how the past has informed our assumptions. It will clarify the meaning of the work that has gone before and give us insight into where the research is going, based on trends developed over the past 50 years. It will show us whether the research done has really answered the question of what music perception really is, or only that something unnamed is being measured.

INTRODUCTION

In my Qualifying Paper I described the psychological theories which influenced the work of Carl Seashore, who developed the first test of music aptitude to be widely used in the U.S.. I described the controversy in music aptitude research that broke out in the 1930's (known as the Atomist-Gestalt controversy). In this dissertation I will show how the psychological theories that underlay the Atomist-Gestalt controversy of the 30's have evolved and influenced research in music aptitude. And I will show how research in music aptitude has changed due to the development of new theories in the field of psychology regarding how we perceive and discriminate.

The Research Question

In this dissertation I will describe and compare the different approaches to music aptitude research and the theories that these approaches are based on. Through looking at the evolution of these theories my goal is to help myself and other researchers gain a broader perspective about their own work and thinking, and about the nature of music aptitude, which in turn, may shed some further light on the nature of intelligence and learning.

To this end, I have formulated the following research questions: How have notions about what music aptitude is and how to go about researching it changed, what theories have influenced that change, and what does that tell us about what we are doing now and where to go next?

Justification for Research Question

This historical review and analysis of music aptitude research will help account for the directions taken in the fields of psychology and education as they pertain to learning in the arts and may point the way for the development of new perspectives and ideas for further research. It will also clarify for researchers in the field of music aptitude testing past works, and how these works have helped formulate assumptions in their present work. It may also give insight to where music aptitude research is going, based on trends developed over the past 60 years. This dissertation could tell us how notions of what music aptitude is have changed over time and whether the research done has really answered the question of what music aptitude actually is, or only that something unnamed is being measured. And finally, this dissertation could be instrumental in giving researchers an overview of the movement of ideas and insights at work in the field of music aptitude research today, which in turn can help them generate more grounded, informed and effective research and theory.

Definition of Terms

Music Aptitude:

There exist many different synonyms used in the field for aptitude: talent, traits, abilities, or musicality, but they all seem to be referring to one or another slant on the phenomenon of aptitude. Carl E. Seashore, a pioneer in the field of music aptitude research and testing, originally defined music aptitude in the 1930's as "the ability to learn, or the level of achievement attained, in any one of a number of given categories such as pitch, rhythm, loudness, tonal imagery or melodic imagery" [1]. Mursell, Seashore's most famous critic, defined aptitude in 1947 as "something in the mental organization that makes one good at clerical work, or mechanical work," etc..[2]

The term aptitude, in relation to psychological tests, came out of the "nature-nurture" controversy in the 1920's (and later the 1940's and 60's) when scientists began to divide into hereditarian and

---

[1]. Seashore, Carl E.,Two Types of Attitude Toward the Evaluation of Musical Talent,Psychology of Music, McGraw-Hill,1938 Appendix, p.383

[2]. Mursell, J.L. Psychological Testing, Longmans, Green & Co. N.Y. 1947 p.211

environmental camps. In the process of trying to discriminate which characteristics of intelligence are innate and which are acquired, aptitude came to denote, over a period of time, innate intelligence as opposed to achievement which denoted instead acquired intelligence.[3]

Pitch Perception or Pitch Discrimination:

The concept of sensory discrimination came from the sensory oriented psychological experiments of Wundt and Helmholtz who tried to limit factors in an experimental situation to those aspects of sensation and perception which they (each in his own way,) felt could be more readily controlled and repeated.[4] Following in their theoretical footsteps, Seashore defined pitch discrimination in the following way:

> "Since pitch is the fundamental character of a tone, and pitch discrimination is a measure of the capacity of this sense, it ordinarily may be regarded as the most basic measure of musical capacity that we have."[5].

The philosopher Thomas Reid pointed out in the 18th century a difference between perception and sensation saying in essence that our ability to feel generates

---

3. Psychology in America, Hilgard, E.R., Harcourt, Brace, Jovanovich, 1987 Chap. 13, p.470

4. Hilgard, 1987 op.cit. p. 44

5. Psychology of Music, Seashore, C.E., McGraw-Hill, N.Y. 1938, p.63

sensations, but the outer world is perceived by way of the senses. Berkeley developed the philosophical position that all reality depends upon perception. Early experiments were done in the area of visual perception by Lotze (1852) and Helmholtz, and then later by the Gestalt psychologist Wertheimer (1912)[6].

The development of research in perception includes auditory perception as in Von Bekesy's traveling wave theory, (1947), issues of esthetics, musical perception and specifically perception of pitch. Mursell, a gestalt psychologist, defined pitch as "that aspect of a tone to which we attend when we sing a given tone or say that a tone is a given interval from another".[7]

Melodic Phrases:

Melodic Phrase is not another way of saying melodic contour. Melodic contour is defined by Dowling as a "set of directional relationships between successive tones in a melody".[8] A melodic phrase, on the other hand, denotes more than just the direction of tones; it implies the larger context of formal musical

---

6. Hilgard, E.R. 1987 op. cit. p.133-135.

7. The Psychology of Music, Mursell, J.L., W.W. Norton & Co. N.Y. 1937 p.71

8. Dowling, W.J. & Fujitani, D.S., Contour, Interval and Pitch Recognition in Memory for Melodies.The Journal of the Acoustical Society of America, Vol. 49, No. 2 (2) 1971 p.524

properties with which we interpret the musical meaning of a beginning, middle, and end of the succession of tones. Although melody can never be exactly like a spoken sentence, a phrase has a rhythm.[9] In terms of Massaro, Kallman & Kelly's experiment, a melodic phrase has tone chroma as well as melodic contour.[10] It is in that sense relational. It has structural properties. Researchers in psychology may speak of melodies as tonal sequences, but when analyzing a melody from the view of formal musical properties like cadence, tonality, progression, meter, down beat, etc... a melodic phrase is much more than a tonal sequence. More musical definitions are on p.109 of this paper.

Atomist-Gestalt:

The atomist point of view was also known as the elemental viewpoint, the bottom-up approach[11] or, as Seashore put it, a theory of specifics. Described by Wing as:

> "...those which have attempted to analyze music into its most elementary basic constituents and then

---

[9]. Bernstein, L. The Unanswered Question, Harvard University Press, 1976, 53-115

[10]. Tone chroma refers to the position of a note within an octave. Massaro, D.W., Kallman,H.J. & Kelly, J.L., The Role of Tone Height, Melodic Contour and Tone Chroma in Melody Recognition, Journal of Experimental Psychology: Human Learning and Memory, 1980, Vol.6, No.1, 77-90

[11]. The Mind's New Science, Gardner, H. Basic Books Inc.1985, p.112

to build up tests of a sensory type which aim at assessing these elementary constituents in their most exact form."[12]

The Gestalt point of view in music was represented by Revesz, Mursell, and Wing. It was the view of the primacy of wholes over parts. The Gestalt psychologists eschewed the sensorial approach in music aptitude testing. As Wing put it:

"Clearly both musical ability and musical appreciation are qualities of the whole mind; though they involve auditory discrimination they do not depend solely on the ear."[13]

The Gestaltist's criticisms of atomistic tests like Seashore's were published in music education journals of the day and became known as the Atomist-Gestalt Controversy. More information about the Atomist-Gestalt Controversy is included in the Literary Review section of this paper.

Conservation:

The theory of conservation in developmental psychology is usually used in it's relation to the work of Jean Piaget. Piaget postulated four main stages of psychological development in children: sensorimotor intelligence; intuitive/symbolic thought;

---

12. Tests of Musical Ability and Appreciation, Wing, H. Cambridge University Press, Cambridge, G.B. 1948, p.8

13. Wing, H. 1948 op. cit. p.3

concrete/operational thought; and formal operations. At the concrete/operational stage of development, a child can represent objects from different perspectives in his mind. He can reverse the operations of a given object or objects and can appreciate that a given object can be more than one thing. For example, shown a picture of two blue birds and three black birds the child can see that there are, in fact, five birds in the picture and that they are also five animals that fly, part of a larger group of mammals called birds which come in many different sizes and shapes. The ability to 'conserve' the concept of 'bird' as being constant throughout the manipulation of images is what Piaget meant by conservation. Piaget studied the development of children through experiments in conservation of number, space, time, matter, etc... Conservation in musical learning is defined by Pflederer and Sechrest as:

> "The ability of an individual to retain the invariant quality, i.e. the sameness of a complex musical stimulus in spite of variations in it's presentation".[14]

Cognitive Science:

The Scottish faculty psychology developed three

---

14. Conservation-Type Responses of Children to Musical Stimuli, Pflederer, M. Sechrest, L., Council for Research in Music Education Bulletin, 13, 19-36, 1968.

classes of psychological functions: cognition, a thought; conation, an impulse or action related to the thought; and affection, the resolution of the action which arouses feeling. The word cognition usually refers to thought. Cognitive science came about as a backlash to Behaviorism which many considered had gone about as far as it could go in achieving objectivity without really shedding any new light on the nature of the human mind. In the early 1950's many different fields of research came together to form a new paradigm called cognitive science. The fields which came together were as diverse as cybernetics and psycholinguistics. They included: Wiener's work on cybernetics[15]; the idea of thoughts as being units of information which we now call bits (binary digits) proposed by Shannon[16] in the same year; research in neurology which came to see the human brain as being like a very powerful computer[17]; the linguistic theories of Noam Chomsky and mathematical theories of Marvin Minsky and in anthropology, the feedback theories of Gregory Bateson. All of these theories

---

15. Cybernetics, Wiener, N. Wiley, N.Y. 1948

16. A Mathematical Model of Communication.Shannon, C.E., Bell System Technical Journal, 1948, 27, 379-423; 623-656.

17. "A Logical Calculus of the Ideas Immanent in Nervous Activity" McCulloch, W. Pitts, W. 1943 Bulletin of Mathematical Biophysics, 5, 115-33.

plus many other applications of these ideas came together over time in the synthesis we now call cognitive science.

Organization of the Paper

By creating a historical account of the research done and the theories underlying the research, I will then be able to do a comparative analysis of the lineages and evolution of the theories the research is about. Specifically, the dissertation includes:

1) A description of four of the most influential tests developed in the thirties and forties. Namely, the Seashore Measures of Musical Talent; the Kwalwasser-Dykema Music Tests; the Drake Musical Memory Test and the Wing Standardized Tests of Musical Intelligence.

2) A review of some of the theories which influenced the design of the music aptitude tests which came out in the thirties and forty's and an examination of some of the features of those tests. For example, how they tried to measure pitch. In addition, I have included a description of the research that has taken place over the last twenty years in music aptitude. Specifically focusing on the work of Deutsch, Bamberger, Attneave & Olson, Dowling, Gardner & Davidson, Pflederer & Sechrest, Massaro, Kallman

& Kelly and others.[18]

3) A look at some of the follow-up studies which were done as a result of the first batch of music aptitude tests and what conclusions they came to about music aptitude tests, what criticisms they had for the tests.

4) an analysis of some features of the four tests named above, specifically the pitch subtests, and a comparison of their designs to some of the newer research in pitch perception to see what differences and similarities they might share. From comparing and examining the designs of the pitch subtests, I will be able to see what researchers today have learned about music aptitude as opposed to what the early aptitude testers were trying to do.

5) Finally, a look at the work of Jeanne Bamberger as an example of an alternative approach to research design in music perception and an analysis of her work in comparison to the work of Diana Deutsch, who exemplifies a more traditional approach to pitch perception research. I will then present an analysis of the results of a small informal study of pitch perception that I did with my students last fall that was based on

---

18. See Bibliography for list of works studied.

one of Deutsch's experiments. I will propose some alternative ways of thinking about the design of experiments in music perception based on my own experiences as a musician and teacher and in light of information gained from doing this historical review.

6) Returning to my research questions, my conclusions and questions for further research comprise the final chapter of this dissertation.

DESCRIPTION OF THE TESTS

This chapter describes the elements of each of four tests of musical aptitude. Each of these tests was developed during the heyday of aptitude testing in the 1930's and 40's. At the end of the descriptions there is a chart to illustrate the similarities and differences of the tests in their approach to measuring musical aptitude and some analysis of the theoretical background of each test.

The Seashore Measures of Musical Talent:

Seashore first published his Measures of Musical Talent in 1919.[1] Originally the tests consisted of 5 subtests: sense of pitch, sense of intensity, sense of time, sense of consonance and tonal memory. Later the consonance test was dropped due to criticisms (Mursell,1931) and a sense of rhythm test was added. Seashore described his Measures of Musical Talent this way in 1938:

> "In general, we may say that in all cases the first step should be the measurement of basic capacities, each of which measures receptivity for one of the four musical avenues, namely, the tonal, the dynamic, the temporal, and the qualitative. In addition to these, immediate memory, a sense of consonance, tonal imagery, and intelligence should be measured. The instrumental needs for this battery

---

[1] Seashore, Carl E. Manual of Instructions and Interpretations of Measures of Musical Talent, C.H. Stoelting, Chicago, 1919.

are met for the purpose of group testing by the Seashore Measures of Musical Talent."[2]

Farnsworth describes Seashore's philosophy about the Measurements in the following way:

"Seashore argued that since music is in essence a matter of pitch, intensity, time, memory, consonance and rhythm, discrimination tests in these areas should make it possible to pick out the potentially musical, with those having the best acuities being expected to give the greatest musical promise. Seashore believed that his tests tapped basic physiological capacities which were inborn and could not be influenced by training. He admitted that his test battery was limited, that there were other capacities he was not measuring.[3]

In a later version of the Manual of Instructions (1960) the tests are described in this way:

"...Not all of the facets of musical aptitude are known, but there are several fundamental capacities that can be assessed.

The Seashore Measures of Musical Talents provide separate measures for six of these capacities: pitch, loudness, rhythm, time, timbre, and tonal memory."[4]

Notice that the description of the test as measuring "fundamental capacities" has not changed. This is particularly striking in light of the recent work of

---

[2]. Seashore, Carl E. 1938 op.cit.

[3]. Farnsworth, P. The Social Psychology of Music, Iowa State University Press, 1969

. [4] Seashore, Carl E. Measures of Musical Talents Manual, (revised) The Psychological Corporation, New York, 1960

Gardner and his theory of multiple intelligences[5] in which he talks about the reemergence of interest in the nineteenth century "faculty psychology" from which Seashore's theories came.

In the pitch test, 50 pairs of individual tones are presented. The subject is asked to judge whether the second tone sounds higher or lower than the first. In the loudness test, the subject is asked whether the second is stronger or weaker than the first. In the rhythm test, the subject is to indicate "whether the two patterns in each pair are the same or different". (from the Manual of Instructions and Interpretations for the Seashore Measures of Musical Talents, 1939 revision). The test of the sense of time consists of 50 pairs of tones of different durations and the subject again decides whether the second pair is longer or shorter than the first. The test for timbre is different in description:

> "The purpose of the timbre test is to measure ability to discriminate between complex sounds which differ only in harmonic structure. It consists of 50 pairs of tones; in each pair the subject is to judge whether the tones are the same or different in timbre or tone quality. The tones were produced with a special generator. Each tone is made up of a fundamental component, whose frequency is 180 cycles, and its first five overtones. Tonal structure is varied by reciprocal alteration in the intensities of

---

5. Gardner, H. Frames of Mind, The Theory of Multiple Intelligences, 1983, Basic Books,NY p.284 2nd.pp.

the third and fourth harmonics." [6]

But the timbre test is basically the same in execution, 50 pairs of tones. The tones are the same pitch but they don't sound alike. The test for tonal memory consists of 30 pairs of tonal sequences. Each pair has one note different in the two sequences. The subject is to indicate which note is different by number.

Many small changes have taken place between the first publication of the tests in 1919 and the present version which I have described above. The test for loudness was originally called a test for intensity, the time test was originally unfilled time, now it is durations of tones. Also, the original battery was comprised of one level of tests with 100 or 50 pairs of tones. Now there are two different levels of difficulty, an easier "dragnet"[7] type test and a more advanced test. The number of pairs of tones in each subtest has been cut down to 50 (or 30, depending on the test).Seashore also changed the title of the tests from Measures of Musical Talent to Measures of Musical

---

[6]. Seashore, 1960 op.cit.

[7]. A "dragnet" type test is supposed to cover a broad range of possible subjects and come to broad conclusions. Seashore does not define "dragnet" in his use of it, so I am supplying this definition in its general use.

Talents in his 1939 edition of the tests. Farnsworth says in a review printed in the Mental Measurements Yearbook for 1939:

> "Seashore's notion that musicality is a matter of multiple capacities is emphasized in the directions for his new battery by a change in titling from "talent" to "talents".

Kwalwasser-Dykema Music Tests:

The Kwalwasser-Dykema test was published in 1930.[8] It was designed along the same general lines as the Seashore test, but there are specific differences. The tests consisted of tonal memory, made up of 25 pairs of melodies. Each melody consists of four to nine notes. One of the melodies is played on the piano, then it is played again. The second playing is either the same or different from the first. The subjects judge whether the melodies are the same or different. The pitch test is different from Seashore's in an intriguing way; A single tone is heard for three seconds. If the tone rises or falls in pitch and then returns to the original pitch it is different. If the tone stays the same without any fluctuations it is the same. There are forty items in the test.[9] In the

---

8. Kwalwasser,J. Dykema,P.Kwalwasser-Dykema Music Tests. Carl Fischer Inc. 1930

9. An Objective Psychology of Music, Lundin, R.W. The Ronald Press, N.Y. 1953 p.248

intensity test, thirty tones and chords are repeated at different degrees of loudness. The subjects judge whether the second tone is stronger or weaker than the first. In the time discrimination test, there are twenty-five items of three tones each. The two outer tones remain the same while the tone in the middle is either of the same duration as the two outer tones or different duration. The subject judges which middle tones are the same or different than the outer ones.[10] The rhythm discrimination test is described as follows:

> "The rhythm discrimination test is a measure of the ability to detect the presence or absence of variations of time or loudness in a rhythmic pattern. The test is made up of 25 paired rhythmic patterns, in some of which the second pattern is identical with the first, in some of which there occurs a change in time, and in some a change in loudness or accent. The patterns consist of two pitches, the last tone of each pattern being C' and the other tones a half-step below. The trials increase in difficulty in the number of tones they contain and in complexity of the rhythmic pattern. The judgement to be made is whether the paired rhythms are the same or different."[11]

The quality discrimination test is a test of timbre discrimination. The test has thirty items. Each item is a pair of notes played on one instrument followed by the same two notes played on the same or a different instrument. Subjects judge whether they are the same or different. Included in the K-D tests are

---

[10]. The Psychology of Music, Schoen, M. Ronald Press, N.Y. 1940, p.178

[11]. Schoen, M. 1940 op. cit. pg. 178-179.

tests for tonal movement, melodic taste, pitch imagery and rhythmic imagery. Whether these four tests are aptitude tests is questionable. The melodic taste test calls for aesthetic judgments based on which of two endings of a melody are more appropriate. The tonal movement test is similar in that it asks for a judgement whether a given tone at the end of a melody gives the best possible resolution of the melody.[12] The pitch and rhythm imagery requires musical training to answer correctly. As a result, they are more of an achievement test so I have left them out of my charts and analysis at the end of this chapter. The subject must compare the tones notated on a printed page to what he/she is hearing and judge whether they are the same or different.

Follow up studies were done for success at the tests by siblings (Swift,1940).[13] Low reliability coefficients were reported by Bienstock(1942).[14] The test was revised with weighted scoring keys and new

---

12. Lundin, R.W., 1953 op. cit. p.249

13. Swift, F.F. A Correlation of Kwalwasser-Dykema Test Scores Earned by Siblings. Unpublished thesis, Syracuse University, 1940. Pp. 157

14. Bienstock, S.F.,'A Predictive Study of Musical Achievement' Journal of Genetic Psychology, 61,135-45 1942

norms by Holmes(1954)[15], which improved their reliability a bit but their validity is still in question. This test is now out of print.

The Drake Music Tests:

The Drake Music Tests consisted originally of four tests. The musical memory test had twenty-four two bar original melodies with four possible variations: change of key; change of time; change of note; or the same melody repeated. The subjects note S for same melody, K for change of key, T for change of time, or N for change of notes. The number of variables increases as the test goes on, making it increasingly difficult. The test is given on a piano. In the interval discrimination test two musical intervals are played on a piano. The subject is to determine whether the last interval is longer or shorter than the first. The first interval is played on lower notes than the second interval so that there will be no overlapping notes. The first forty intervals are always ascending the last forty are in exact reverse order of the first forty and are always descending. Drake concluded from his results that the smaller the ratio between intervals, the more difficult the

---

[15]. Holmes, J.,'Increased reliabilities, new keys and norms for a modified Kwalwasser-Dykema test of musical aptitudes', Journal of Genetic Psychology, 85, 65-73, 1954

discrimination between them. Drake developed the retentivity test as a test for memory of isolated tones or absolute pitch. There are twenty trials in the test. Each trial is made up of two sections. In each section are three items to be remembered and compared with three items in the other section. To illustrate: first you hear an interval of two notes, then you hear a metronome beating at 164 beats per minute; then you would hear three notes given. That would comprise the first section. A lapse of time occurs. (Drake does not say how long) Then you would hear the second section. First you would hear another two tone interval. The subject is to determine whether it is longer or shorter than the first two tone interval in the first section. Then the subject would hear a rate of time given on the metronome of 78 beats per minute. The subject has to determine if it is faster or slower than the rate of the metronome in the first section. Finally, one note is given. The subject has to determine whether the note is: 0= none of the three notes in the first section, 1= the first of the three notes in the first section, 2= the second of the three notes in the first section or, 3= the third of the three notes in the first section; The intuition test is more like a test of aesthetic judgement even though Drake claims that this kind of judgement is innate. The test is made up of three subtests of 24 trials each. Each trial consists of two musical phrases. One group of trials is supposed to determine the subjects ability to discriminate correct "phrase-balance". Meaning that the answering phrase is the same length as the first phrase. One group of trials is supposed to determine the subjects ability to discriminate "key-centre". Whether the phrase returns to the tonic of the first phrase. The final group is supposed to show whether the subject can discriminate "time-balance". Whether the answering phrase should end in the same time as that in which the first began. The subjects have only to mark whether the answering phrases are "right or wrong". The subjects were told which of the three subtests they were being given and the subtests were not mixed.

Of the four original tests, only musical memory and interval discrimination were found by Drake to be useful (with respect to their reliability and validity) (Drake, 1933).[16] He made a recording of the musical

---

[16]. Drake, R.M. 'Four new tests of musical talent' Journal of Applied Psychology, 17, 136-47 1933

Drake, R.M. 'The validity and reliability of tests of musical talent' Journal of Applied Psychology, 17, 447-458 1933 memory test in 1942 and a rhythm test in 1954.[17] Lundin (1949) found the validity figures to be lower than Drake had predicted.[18] Gordon studied the effects on training and practice (1961) and Ferrell did a follow up study in which he found that the test successfully discriminated between students with a high level of musical aptitude and those without.[19]

Wing Standardized Tests of Musical Intelligence:

The Wing Standardized Tests of Musical Intelligence (1948), consisted of seven tests. The pitch test contains thirty pairs of chords. In some of the pairs one note is either higher or lower. In other pairs the chords may change from major to minor or the two chords are the same. The subject marks on a printed sheet U for up, D for down or S for same. In

---

17. Drake, R.M. Manual for the Drake Musical Aptitude Tests, Science Research Associates,Chicago, 1954-57

18. Lundin, R.W., 'The development and validation of a set of musical ability tests' Psychological Monographs, 63:305, I-20 1949

19. Gordon, E., 'A study to determine the effects of training and practice on Drake Musical Aptitude Tests scores',Journal of Research in Music Education,9 (I), 63-74,1961

Ferrell,J.W., A validity investigation of the Drake Musical Aptitude Tests, PhD Thesis, Music, State University of Iowa. 1961 the memory test thirty pairs of melodies range in length from 3 to 10 notes. One note of the second half of each pair is altered. The subject indicates the number of the altered note. There are fourteen items in the rhythmic accent test. Subjects hear two versions of a short melody and judge whether they prefer the first or second melody and whether they are the same or different. The harmony, intensity and phrasing tests have the same form as the rhythmic accent test. The chord analysis test consisted of 20 items in which the subject was asked to count the number of pitches present in the chord.[20]

Follow up studies were done by Bentley (1955), a comparative study, by Heller (1962) to study the effects of formal music training on Wing test scores. More studies comparing the Wing tests to other tests were done in the 1960's, Cain (1960) did a comparison study of the Wing test to the Gaston and Drake tests. Heim (1963) studied the musical aptitude of blind subjects using the Wing test and Wertz (1963) studied the relation between changes in musical preference and scores on the Wing test.[21]

---

[20]. Tests of Musical Aptitude, Lehman, Prentice-Hall, New Jersey, 1968 p.47

[21]. Bentley's dissertation was very valuable for me to read because his interests were similar to mine but from an earlier perspective. Bentley,R.R., A critical comparison of certain aspects of musical aptitude tests. PhD. thesis,

Follow-up studies on all four of these tests will be discussed in more detail in the chapter of this dissertation called Follow-up Studies. What follows here is a chart to illustrate the differences and similarities of these four tests and some analysis of the theoretical background of the tests. In the chapter on Pitch Subtests of this paper I will examine these four different approaches to testing pitch perception specifically and look at some more recent experiments in pitch perception to compare what researchers today have learned about pitch perception as opposed to what these early aptitude testers were trying to do.

Test Comparisons

The Test Comparisons chart has eight vertical

University of Southern California. 1955

Heller, J.J., The Effects of Formal Music Training on the Wing Musical Intelligence Scores. Doctor's thesis, State University of Iowa, (Iowa City, Iowa) 1962 (DA 23:2936)

Cain,M.L., A Comparison of the Wing Standardized Tests of Musical Intelligence with a Test of Musicality by Gaston and the Drake Musical Aptitude Tests. Master's thesis, University of Kansas (Lawrence, Kansas) 1960.

Heim,K.E., Musical Aptitude of Senior High School Students in Residential Schools for the Blind as Measured by the Wing Standardized Tests of Musical Intelligence. Master's thesis, University of Kansas (Lawrence, Kansas), 1963.

Wertz, C.B.,Relation of Changes in Musical Preference to Scores on the Wing Standardized Tests of Musical Ability and Appreciation. Master's thesis, University of Kansas, (Lawrence, Kansas). 1963

Tester:

TEST COMPARISONS

Test							
Seashore Measures of Musical Talents	Subtest Title:	Pitch	Loudness	Rhythm	Time	Timbre	Tonal Memory
	Stimulus:	Pairs of Tones	Pairs of Tones	Pairs of Rhythmic Patterns	Pairs of Tones of Dif. Duration	Pairs of Tones	Pairs of Melodies
	Subject Response:	Higher/Lower	Stronger/Weaker	Same/Different	Longer/Shorter	Same/Different	Which note is Different by #
Kwalwasser-Dykema Music Tests	Subtest Title:	Pitch	Intensity	Rhythm Discrimination	Time *	Quality *	Tonal Memory
	Stimulus:	Single Tone (which wavers)	Pairs of Tones	Pairs of Rhythmic Patterns of Different Durations	3 Tones: 2 Outer Ones Stay the Same, Inner Changes Duration.	Pairs of Tones (played on different instruments)	Pairs of Melodies
	Subject Response:	Same/Different	Stronger/Weaker	Same/Different	Same/Different	Same/Different	Same/Different
The Drake Music Tests	Subtest Title:	1)Retentivity 2)Interval *Discrimination*	Retentivity 1) 2 note *interval* 2) Metronome 3) 3 notes	Intuition Pairs of *Melodies* (phrase-balance)	Intuition Pairs of Melodies (time-balance)		1)Intuition 2)Musical Memory
	Stimulus:	1) (see next column) 2) Pairs of Tones	1) same 2) same 3) 1 note of 3				1)Prs. of Melodies 2) 24 Melodies, 4 Variations- Key, Time, Note or Same
	Subject Response:	1) (see next column) 2)Long/Short	1)Long/Short 2)Fast/Slow 3) 1st,2nd,3rd note or none.	Right/Wrong	Right/Wrong		1) Key Centre-Right/Wrong 2)K=key,T=time N=note,S=same.
Wing Standardized Tests of Musical Intelligence	Subtest Title:	1) Pitch 2) Chord Analysis	Intensity	1) Phrasing 2) Rhythm			1) Memory 2) Harmony
	Stimulus:	1) Pairs of Chords- in some, 1 note is: Higher/Lower, Major/Minor or Same. 2) Count # of Pitches Present in a Chord.	Pairs of Melodies-	Pairs of Melodies-			1) Pairs of Melodies- 1 note Altered in 2nd. Half. 2) Pairs of Melodies-
	Subject Response:	1)Higher/Lower, Major/Minor or Same. 2)#of Pitches	Like/Dislike & Same/Diff.	Like/Dislike & Same/Diff. (for both 1&2)			1) # of Altered Note 2)Like/Dislike & Same/Diff.

columns and twelve horizontal sections. Within that framework, the four tests compared are broken up into three sections for each test. After the first and second columns, the remaining columns list the titles of the subtests in each test. In the first column is the name of the designer of the test. In the second is a descriptive title for the contents of each section. For instance, in the first section of the second column it says "Subtest Title:" That tells you that the following information in that horizontal section is a listing for the titles of each subtest in that particular test. In the second section of the second column you will see the word "Stimulus:". That tells you that what follows in the second horizontal section is a short description of the stimulus used in each of the subtests listed above in that column. For instance, under Rhythm (in the Seashore subtest) you would find that the stimuli were "pairs of rhythmic patterns". In the third section you will see the words "Subject Response". That tells you that the information that follows in that section and across the columns is a short description of the kind of response that is expected from the subject in the subtest at the top of the column in the subtest title section. This pattern repeats for each test compared. The reader gets an overview of the kinds of subtests, stimulus and responses for each of the four tests compared. Where there are numbers with a parenthesis following, that indicates that two subtests of a particular test both measure aspects of pitch (for instance) or tonal memory. In that sense, all of the subtest title sections can be compared to the subtest title section for Seashore at the top of the table. Where there are numbers with a parenthesis following in the "stimulus" and "response" sections, those numbers refer to which of the above subtests they are describing.

You will notice that the subtest titles of the Drake and Wing tests as opposed to the Seashore and K-D tests have very little in common. I am not implying in my chart that phrasing is the same thing as rhythm. I didn't have room to reflect the different theoretical backgrounds between the first two tests and the second two. What this chart does do is give an overview of what the categories of stimulus and response are in each of the subtests and where they are in relation to each other. Also, as you can see, describing the Drake Retentivity test in a simplified columnar form is a bit too brief. If one hasn't read the description of the test previously, it might seem a little obtuse.

In general, it has been noted that factors of experience are much easier to control in the Seashore and K-D tests than they are in the Drake and Wing tests.[22] But if you look at what is being measured in each instance as opposed to how the measurement is designed, the comparison of "factors of experience" falls apart. The reason for this is that there is no proof that Seashore's and Kwalwasser-Dykema's tests are measuring musical intelligence as opposed to hearing discrimination.[23] Since what the Drake and Wing tests purport to be measuring is musical intelligence, and their methods, as byzantine as some of them seem to be, contain more musical stimuli than the Seashore or K-D tests, it seems to me that (in terms of "factors of experience") there is really nothing to compare. The tests are designed to measure completely different things.

The Comparison of Stimuli and Comparison of Responses charts are modeled after the Test Comparison chart but break out first the stimuli of each subtest to compare them, then the responses of each subtest to compare. When you look at the Comparison of Stimuli chart, you can see more of the theoretical bent of the different testers come out. The kinds of stimuli that they each used was based on their theoretical orientation. For instance, the majority of Seashore's

---

22. Holstrom, L.G., 1963 op. cit. p.174

23. The Psychology of Music, Mursell, J.L., W.W. Norton & Co. New York, 1937, p. 299

COMPARISON OF STUMULI

Tester:	Pairs of Melodies	Pairs of Tones	Rhythmic Patterns	Pairs of Chords
Seashore Measures of Musical Talents	Tonal Memory	Timbre Time Loudness Pitch	Rhythm	
Kwalwasser-Dykema Music Tests	Tonal Memory	Intensity Pitch(1 tone) Timbre	Rhythm Time	
The Drake Music Tests	Retentivity Intuition Musical Memory	Retentivity Interval Discrimination	Retentivity	
Wing Standardized Tests of Musical Intelligence	1) Phrasing 2) Rhythm Intensity Harmony Memory			Pitch Chord Analysis

subtests used Pairs of Tones as the stimuli, which make sense considering his atomistic and psychoacoustic orientation. On the other hand, almost all of Wings tests are in the Pairs of Melodies column, reflecting his more Gestalt approach to testing. One interesting stimulus is the Kwalwasser-Dykema Pitch test stimulus: one tone held for three seconds which wavers either up or down, or stays the same. This is fascinating to me because I "hear" music, especially singing, as moving in just this way. The fact that they designed their pitch test this way doesn't seem to fit with their sensory/elementalist theoretical background on the surface of it.

In the Comparison of Responses chart the column marked Right/Wrong,Like/Dislike is the part of those tests of Drake and Wing which have been called into question as not really belonging in an aptitude test. They call for judgments based on aesthetics and achievement rather than aptitude. Drake says of his Intuition subtest:

> "According to the theory upon which these principles are built, (Gestalt) no explanation of what constitutes a good or a bad, or "right" or "wrong," answering phrase is necessary, for some, who are musical, will feel the difference intuitively and those who are not musical will not discriminate the difference."[24]

---

[24]. Tests of Musical Talent, Drake, R.M., Journal of Applied Psychology, 1933, p.146-147

COMPARISON OF RESPONSES

Tester:	High/Low	Strong/Weak	Faster/Slower	Long/Short	Same/Different	Right/Wrong Like/Dislike	Maj./Min.	# of Factor	# of Altered Note/Pitch	# of Pitches
Seashore Measures of Musical Talents	Pitch	Loudness		Time	Rhythm Timbre				Tonal Memory	
Kwalwasser-Dykema Music Tests		Intensity			Pitch Rhythm Discr. Time Discr. Quality Discr. Tonal Memory					
The Drake Music Tests			Retentivity	Retentivity Interval Discr.		Intuition (time bal.) (Key Centre) (phrase bal.)		Musical Memory (K,T,N or S)	Retentivity	
Wing Standardized Tests of Musical Intelligence	Pitch				Intensity Phrasing Rhythm Harmony	Intensity Phrasing Rhythm Harmony	Pitch		Memory	Chord Analysis

There is no more evidence of this point of view being correct than there is that the Seashore tests measure anything other than hearing discrimination. A strange factor (to me) of the Wing subtest called "Chord Analysis" is that the subject is asked to "count the number of pitches in a chord" rather than comparing two chords in which one note has been altered. This only strikes me as strange because Wing was characterized as an "Omnibus-type" tester[25] and I would have thought he would be interested in the figure/ground aspects of perceiving a chord. This design doesn't seem like it would address that issue at all. This design would tell the tester whether the subject was discriminating notes in a chord but not necessarily which notes, i.e. the tonic, the third, the fifth, etc..

In summary, the Seashore and Kwalwasser-Dykema tests are designed to measure "capacities" and for the most part conform to the elementaristic views of their designers. The Drake and Wing tests are designed with the basic premise in mind that musical perception is a function of the mind, not the ear. Further, the tests reflect the Gestalt principles of proximity and good continuation in the use of melodies as a basis for

---

[25]. Lehman, 1968, op. cit. p.48

perception in music. In looking at the work that came after these tests in the field of music aptitude research, the use of melodies as the basis of experiments in music perception predominates. Thus, I would assume that there is some real value in what the Gestalt-oriented early testers were trying to do. I will compare and analyze the pitch subtests of these four tests in relation to more recent experiments in pitch perception in the chapter entitled "Pitch Subtests".

LITERATURE REVIEW

Carl Seashore developed the first psychological measurement of musical ability in the United States. Based on the sensorial methods of psychological experimentation in use at the time, the test: Seashore Measures of Musical Talent (C.H. Stoelting, Chicago, 1919) purported to measure "... six capacities or abilities for the hearing of musical tones"[1].

Carl Emil Seashore:

Carl Emil Seashore was born in 1866 in the United States. His parents were from Sweden and he was brought up in a strict Lutheran family. He went to Yale University for his undergraduate studies and helped support himself there by leading a choir to which he would charge his fellows a nominal fee to belong.[2] He spent most of his academic career at the University of Iowa where he published his tests and wrote many books and papers under the auspices of the Studies in Psychology program.

Seashore has made references to many psychologists and researchers in his work including

---

1. Seashore, Carl E., Psychology of Music, McGraw-Hill, 1938, "Two Types of Attitude Toward the Evaluation of Musical Talent". Appendix, p.383

2. Private Conversation with Edwin Gordon, holder of the Carl E. Seashore Chair in Music Education at Temple University, Philadelphia, PA. Nov. 1986.

Fechner, Galton, Helmholtz and William James but the predominant influences in his theories on testing came from Wundt through Titchener and Cattell. Although not of the structuralist/sensationist school, the work of Carl Stumpf in Germany, who worked with a young musical genius and outlined the basic abilities a superior musician should possess,[3] and the work of Revesz and his theories of musical types; creative and reproductive-interpretive,[4] are also cited.

Revesz was a proponent of the Gestalt view of music aptitude. He divided pitch perception into three categories: "universal genuine absolute pitch", by which he meant a person who is able to identify any note in the entire musical range; "limited absolute pitch", those who are able to identify notes only within a given range of pitch; "regional pitch", in which persons identify the tone based on the "tone region" to which it belongs; and finally, "standard pitch", in which a musician or other musical person memorizes a particular tone over a long period of time.

---

[3]. Stumpf, Carl. Akustische Versuche mit Pepito Areola, Zeitschrift fur Experimentelle un Angewandte Psychologie, II (1909), 1-11.

[4]. Revesz, G. <u>The Psychology of a Musical Prodigy</u>, Harcourt, Brace Janovitz, New York, 1925.

Revesz further distinguished between absolute and relative pitch:

> "The first (absolute pitch) is the spontaneous identification and reproduction of notes in isolation; the latter (relative pitch) of intervals."[5]

Revesz is probably best known for his classification of two musical types: creative and reproductive/interpretive. The creative type was made up of composers and composer/musicians. The reproductive type was composed of persons whose talent was for conducting and playing music. Although the creative person was often also a musician, the reproductive/interpretive was rarely a good composer. Revesz saw these two different types of musicality as fundamentally different. In retrospect, these differences may have a cognitive basis but Revesz saw musicality as being an inborn talent. From the modern day perspective, these ideas seem culturally bound, especially the idea that persons who would be identified as reproductive/interpretive in his theory wouldn't be able to create anything musically significant. There is more to the modern world of music than classical or romantic symphonic music. Many performers are also composers, especially in Jazz and in "new music". String quartets write their own

---

[5]. Introduction to the Psychology of Music, Revesz,G. University of Oklahoma Press, 1954 p.108

pieces, combine with elements of folk music, bringing mandolins back into the world of classical music and with them, the composing tradition of bluegrass and acoustic music. This doesn't even begin to describe all of the composing/performing that presently goes on in the fields of popular and world music. Revesz does give this one justification for his position:

> "In olden times, when playing and composing, singing and writing went hand in hand, this combination was found far more frequently than it is today. However, in those days such extraordinary demands were not made of the executant musician as is the case today. The same is still true among musicians of primitive races."

I think Revesz was describing a phenomenon of the western world's "golden age of classical music" not a fundamental characteristic of musicality. How could he have guessed that the music of those so called "primitive races" would be considered on the cutting edge of musical awareness today? On the other hand, it may very well be that there is a different kind of cognitive activity going on when one is composing than when one is interpreting a piece of music, but I don't think it has anything to do with inborn ability or the size of the composition or extraordinary demands on composers.

Revesz had a marked influence on all of the researchers in the field of music aptitude. He is often quoted in relation to his work with musical prodigies and as evidence for the gestalt view of musicality. The two opposing views, the Atomistic view as personified by Seashore and the Gestalt view as personified by James L. Mursell, a Gestalt psychologist at Columbia, came to loggerheads in the 1930's in a series of articles in the Music Educator's Journal. The controversy came to be known in the field of music psychology as the Atomist-Gestalt Controversy.

The Atomist-Gestalt Controversy

The Atomist point of view was personified by Carl Seashore in his defense of his testing procedures:

> "Musical talent is not a single talent: it is a hierarchy of talents, many of which are entirely independent of one another."[6]

The Gestalt point of view came to be associated with the critiques that were written about the Seashore tests, particularly by Mursell. Mursell described the Atomist-Gestalt controversy in his book on psychological testing in 1947:

> "Existing psychometric instruments are said to be based upon the presuppositions of an atomistic or mechanistic psychology. Of necessity they undertake to isolate and measure separate abilities, such as general intelligence, interest, mechanical aptitude, sociability, musical talent, and the like. There seems no other procedure, so far as can be seen at the present time. These abilities are thought of as independent unitary functions, and the individual human mind is, at least by implication, regarded as the sum total of these units which exist in it in ascertainable amounts....

---

[6]. Seashore, Carl E. 1919 p.6, op.cit.

This whole viewpoint, however, it is argued, is erroneous. It is diametrically opposed to the organismic or holistic or configurationalist psychology coming more and more into prominence. The individual mind is precisely not a composite of unitary traits or abilities, but a functioning unit. Intelligence, for instance, cannot be separated from interest. What is called musical talent, or artistic talent, or mechanical aptitude is not a sort of special faculty, but is essentially the mind or personality as a whole operating in a particular way."[7]

Lundin described the opposing points of view in 1985:

"For him, (Seashore) musical talent consisted of a number of separate capacities which were fairly independent of each other... Here we have a perfect psychophysical parallelism. Each function of the mind has a physical counterpart. Wundt and Titchener, the founders of structuralism, would have indeed been proud. Structuralism, as a mentalistic system of psychology, also followed an elemental viewpoint toward conscious experience. Its aim was the analysis of conscious experience into atoms and molecules of feelings and sensations. This is, of course, in direct contrast to a view which considers the mind as an integrated unanalyzable whole.[8]"

Edwin Gordon has delineated the two theories in this way:

"The Gestalt group holds that music aptitude is a unitary trait of which general intelligence is a substantial part. The Atomistic group contends that music aptitude is multidimensional, that it has various parts, and that none is significantly related to general intelligence."[9]

---

7. Mursell, James L. Psychological Testing, Longmans, Green & Co. New York, 1947 p.22

8. Lundin, R.W. An Objective Psychology of Music, R.E. Kreiger, Malabar,Fla. 1985 p.207

9. Gordon, 1984 op.cit.

The Principal Players:

Seashore defined himself as a "specifist" and his detractors as the "Omnibusts" and illustrates the two opposing theories in the following way:

> "Let me designate his (Mursell's) theory as the "omnibus theory" and mine as the "theory of specifics," Somewhat on the analogy of the distinction between cure-alls and specifics in drugs. Since his view was stated specifically in part against my six Measures of Musical Talent, now available on phonograph records, I may simplify my argument in the limited space by speaking only of the issue involved in these six measures.
> 1. They represent the theory of specific measurements insofar as they conform to the two universal scientific sanctions, on the basis of which they were designed; namely, that (1) the factor under consideration must be isolated in order that we may know exactly what it is that we are measuring, and that (2) the conclusion must be limited to the factors under control.
> Each of these six tests purports to measure one of six capacities or abilities for the hearing of musical tones. There is little overlapping in these functions, and their isolation for the purpose of measurement has been criticized only in the case of one.... I deliberately coined the term "measure" for this type of procedure in order to indicate its scientific character and distinguish it from the ordinary omnibus theory procedure in tests."[10]

Mursell, Seashore's chief critic in print and the person whom Seashore saw as the main proponent of the omnibus theory had the following comments to make about the revisions of the Measures of Musical Talents in 1939 which illustrate a few of the points of the

---

[10]. Seashore, C.E., Two Types of Attitude Toward the Evaluation of Musical Talent, Music Educators Journal, Dec. 1937.

Gestalt group:

> "The claim is that pitch discrimination, loudness discrimination etc.. are identifiable and measurable abilities. No one can deny this. But the crucial question is whether they operate in music as they do in the highly isolated test situation. No answer at all based on facts is forthcoming. Apparently one is invited to take it on faith. Yet it seems decidedly improbable. It has been shown again and again in the history of mental testing that isolated sensory and perceptual abilities do not indicate performance in complex situations involving context and meaning. The instrument is a reversion to a type of psychophysical and sensory testing that belongs to the infancy of mental measurement, and has repeatedly been proved worthless as an index to higher mental abilities. That it measures pitch discrimination, loudness discrimination, immediate memory for tonal nonsense-items, etc., is undeniable. That it measures anything that by any stretch of the imagination can be called musical talent is highly questionable and entirely unproven."[11]

Along with Seashore and Mursell in this controversy were a number of other music researchers, psychologists and educators, all of whom had a perspective on music aptitude testing and educational testing in general.[12] However, I will focus on the four principal writers whose articles appeared in the Music Educators Journal in the fall of 1937 and the Spring of 1938: Seashore, Mursell, Jacob Kwalwasser and William S. Larson. Jacob Kwalwasser, designer of the Kwalwasser-Dykema music test and advocate of the hereditarian school of

---

[11]. Mursell, J.L., Review of the 1939 revision of the Seashores Measures of Musical Talents, Mental Measurements Yearbook, 1939

[12]. See Appendix I

musical aptitude was a strong supporter of Seashore's:

"The basic structures that make musical behavior possible are inherited, for we are born with all the equipment that is ours to use. Nor can talent be acquired from the teacher, no matter how eager the student may be for musical success." [13]

and William S. Larson, chairman of the music education department of the Eastman School of Music, who used Seashore's tests every year on all enrolling students for twenty five or so years:

"Ever since these tests appeared in 1919, considerable controversy has existed as to their value in music education. However, this most significant point has been beclouded by arguments on certain theoretical issues such as whether suitable criteria for judging consonance have been advanced. Another issue has been whether the tests should be used as a composite battery with average scores, possibly with some weighted deviation of such, in determining a classification of talent (Professor Seashore opposes this plan which he calls the omnibus theory) or whether each individual test in the battery should be considered a test unit in itself for measuring that specific capacity for the value it in itself can lend to musical prognosis (this theory of specifics Professor Seashore approves).[14]

The Music Educators Journal:

The series of articles which appeared in the

---

[13]. Kwalwasser, J. Exploring the Musical Mind, Coleman Ross, New York, 1955 p.161

[14]. Larson, W.S. Review of the original 1919 edition of the Seashore Measures of Musical Talent, Mental Measurements Yearbook, 1939

Music Educators Journal do not represent the entire scope of the debate which had been raging in music aptitude circles for years, but they do illustrate the main points of the two sides, the Atomists and the Gestaltists. I will present the articles in chronological order and analyze their relative positions as I go along.

What About Music Tests?:

In October of 1937 Mursell published a short paper in the Journal entitled "What About Music Tests?"[15] Judging from the first two paragraphs, Mursell is concerned with what he considers to be rash claims made by researchers such as Seashore, Kwalwasser, Stanton and Larson over the validity of the Seashore tests:

> "We have every right- nay, we have a positive duty- to demand stringent proof that any given test will really do what it promises. And my objection to the existing music tests is very simple. They have never been proved up.
>
> By far the most searching and fundamental question- though by no means the only one- to ask in judging any test is: Does it really measure what it purports to measure? Is it valid? Its title may be just as misleading as the label on a quart bottle of Doc Whoosis' Herbal Remedy, guaranteed to cure rheumatism, paresis, lumbago, and cancer. The woods are full of published tests with dishonest titles. Various persons have assembled a few items and then proceeded to tell the world that they had tests for intelligence, or moral character, or personality, or emotional trends, or what have you. But only those given to extreme credulity will believe them. We

---

15. Mursell, J.L.What About Music Tests? Music Educators Journal,Oct.-Nov. 1937 p.16-18

> think that the Terman Group Test of Intelligence really does measure something- though just what it is we know not- called general intelligence not because Dr. Terman says it does, but because Dr. Terman has developed a proof of his claim and because he has published his proof so that we can study it and form our own judgement. What then about the music tests? After a careful examination of all the research studies I have been able to find, and they are not few, I am compelled to the opinion that in the case both of the Seashore Measures of Musical Talent and the Kwalwasser-Dykema Music Tests such proof is entirely lacking."[16]

Mursell furnishes a table of all the important validation studies done[17] and states the following:

> "This is the standard technique used for proving that an intelligence test, for instance, really measures intelligence. And here we have the research jobs on which it has been applied to the Seashore tests. And what does our table show? There is only one answer, and a sad one. We see very little relationship between rating on the tests and competent "musical behavior". Anger, shouting, and propaganda can no more alter the significance of these figures than they can alter the significance of your bank account."[18]

He gives a sample of the kind of research in which the tests are being used. The example is Kwalwasser's use of an unpublished study by Lenoire to support Kwalwasser's claim that Negroes have better rhythm than Whites:

> "According to Kwalwasser the Negro child "was found far superior to the white child in rhythm..." and this is offered as a scientific tid-bit to the

---

16. Mursell, 1937 op. cit.

17. See Appendix III

18. Mursell, 1937 op. cit.

unsuspecting music supervisor. Well, we may not be scientific experts, but still we can count; so let us ask just how superior was the Negro child? Here we have it! "The colored child averaged 65.69 in rhythm, whereas the white child averaged only 61.48." A difference of 4.21 entire points! Copernicus and Darwin had nothing on this."[19]

Two Types of Attitude Toward the Evaluation of Musical Talent:

Seashore's reply appeared in the next issue of the Journal and in it he addresses directly the nature of the two schools of thought, Atomist and Gestalt, as he sees them. In addition, he goes on to support his position that the tests validate what they purport to measure using his "scientific" methods of experimentation as proof of validation:

"This is an internal validation in terms of success in the isolation of the factor measured and the degree of control of all other factors in the measurement. When we have measured the sense of pitch, that is, pitch discrimination, in the laboratory with high reliability and we know that pitch was isolated from all other factors, no scientist will question but that we have measured pitch. There would be no object in validating against the judgement of even the most competent musician. We would not validate the reading on a thermometer against the judgment of a person sensitive to temperature."[20]

Seashore goes on to make the following points:

---

19. Mursell, 1937 op. cit.

20. Seashore, C.E. Two Types of Attitude Toward the Evaluation of Musical Talent, Music Educators Journal, Dec. 1937

(3) That the relatively low reliability of the data may be caused by the inherent weaknesses of the phonograph record speeds, etc;

(4) that he has always insisted that the tests be used in conjunction with individual case-histories and auditions;

(5) that the tests play the role of a "negative adjustment" in determining whether a child is capable of doing the musical work to which he or she is aspiring and a positive role in the sense of discovering talent amongst a large number of pupils;

(6) the conclusions that are made should be limited to the "legitimate implications of the factors measured", and

(7) the tests have been the victim of propaganda, misuse and particularly misuse by means of the omnibus theory:

> "They have suffered much from popular and superficial advertising and propaganda. I have often paraphrased the aphorism: The Lord protect me from my friends, I can protect myself against my enemies. Among the friends are many who assume a blanket validity of these tests on the omnibus theory and have, therefore, sold the notion on a large scale. This has also been the basis of many journalistic stunts, and there are many wrong applications made. Occasionally my own unguarded statements should have been qualified. This difficulty lies in the fact that nonlaboratory people have been fed up on the omnibus theory."[21]

Seashore has tried to shift the focus of the

---

[21]. Seashore, 1937 op. cit.

controversy off of his tests specifically and on to the relative merits of the two theories of music research. He tries to distance himself from any taint of "unscientific" behavior even to the point of alienating his own group: "The Lord protect me from my friends.."

From the Realm of Guess into the Realm of Reasonable Certainty:

One of the friends he may have been referring to came to Seashore's defense, and his own defense, with an article published in the Music Educators Journal for February, 1938. Jacob Kwalwasser made the following opening remarks in his article,"From the Realm of Guess into the Realm of Reasonable Certainty" (A title which he took from Mursell's book: The Psychology of School Music Teaching, 1931):

> "No serious student interested in the development of scientific procedures for the measurement of musical traits is likely to be misled by the adverse comments comprising the article "What About Music Tests?"...However lest the veneer of scientific verbiage which the author has thrown about his arguments may confuse the less thoughtful, this article cannot be permitted to pass unchallenged".[22]

Kwalwasser takes Mursell's comment that the "woods are full of published tests with dishonest titles" to mean particularly music tests and replies that there are only twenty published music tests

---

22. Kwalwasser, J. From the Realm of Guess into the Realm of Reasonable Certainty, Music Educators Journal, February, 1938 p. 16-17

available. He then goes on to refute the contention of Mursell's that the tests "have never been proved up" by trying to "prove up" Mursell's article:

> "I maintain that his presentation is vulnerable on three counts. (1) Because of inadequate sampling due to (a) insufficiency of numbers and (b) failure to use more heterogeneous sampling; (2) Drawing unwarranted conclusions which go beyond the evidence yielded by experimentation; and (3) The presence of uncontrolled variables in each and every one of the studies reported."[23]

In order to make his point further, Kwalwasser uses an old review of the Seashore Measures which Mursell published in his 1931 book, The Psychology of School Music Teaching, in which Mursell initially seems to praise Seashore's work. After citing the praise Kwalwasser then lets the reader know that the citation was Mursell's own. He tries to attack Mursell's study done in 1932 by citing the works of Starch and Ruch, who did studies indicating the unreliability of teacher's grading methods.

Practical Experience with Music Tests:

William Larson may have had a different motivation for joining the fray; to defend his use of the Seashore tests at the Eastman School and in the public schools of Rochester, New York for so long. He starts out by saying:

> "The faculty of the latter [public schools of Rochester] especially have been somewhat perturbed by

---

[23]. Kwalwasser, 1937 op. cit.

> the matter [of the Seashore tests] and have taken certain criticisms by inference even to be in the nature of an affront to their musical program, for the organization of special music classes of the public school system has come to be associated to a very large extent with the guidance work maintained by their music psychologist. [Dr. Ruth Larson, who worked closely with William Larson at the Eastman School in planning the testing program in the public schools][24]

A good deal of the rest of the article is a personal history of William Larson's career in music education research and a listing of his and Ruth Larson's credentials and qualifications. He finishes up this part of the article inviting the reader to "have the reactions of some of the leading music educators in Rochester appended to this article, so as to allow them an opportunity to state personally their estimates of the values to be derived in this work."[25]

Mursell had said in his article What About Music Tests? that what distinguishes real validity and proof from rash pronouncements was the publishing of results of tests. "...Terman has developed a proof of his claim and has published his proof so that we can study it and test it." The next part of Larson's article seems to address this comment as if Mursell's remarks were specifically made about the Eastman studies done by

---

24. Larson, W. Practical Experience with Music Tests, Music Educators Journal, March, 1938 p. 31,68-74

25. Larson, 1938 op. cit.

Ruth Larson and the Eastman testing program in general:

> "There have at times been questions in regard to the availability of our studies. The situation has not been conducive to formal reports for two reasons:"[26]

His reasons are that he doesn't want to endorse any particular form of testing for musical talent and that he is too busy to make a general presentation of the results of the tests. Larson sums up his article by coming to the following conclusions:

> "To urge teachers of music to minimize the significance of inherent differences in musical capacity is retrogressive; to maintain that success in music depends primarily on the direction of interest and will, and to consider that achievement is the result of a concentration of general ability in the special medium of music are tenets which are contrary to the general observations of teachers of music. And scientific investigation supports this commonly known fact that students differ in talent, and in turn, in achievement- interest and effort being the same, a fact which we often wish were not true and which is apt to make us susceptible to rationalization on our own part or through the writings of others."[27]

The Issues of the Test Discussion:

In the last article, "The Issues of the Test Discussion"[28], Mursell takes the points brought up in Seashore's and Larson's articles and answers five

---

[26]. Larson, 1938 op. cit.

[27]. Larson, 1938 op.cit.

[28]. Mursell, J.L. The Issues of the Test Discussion, Music Educators Journal, May 1938, p.22,23,74-77

questions regarding them. A reply to Kwalwasser he relegates to an appendix. These are the "five important questions" which Mursell feels have been raised in the discussion of the tests. He addresses them in the following way:

> "(1) What further light is thrown on my conclusion?
> (he summarizes Seashore's admissions that correlations between test results and general musical behavior would be low, being the omnibus approach and adds:)
> But Larson seems to feel that evidence for the closer relationship of the tests to talent may indeed exist. He refers to an investigation, evidently very careful, in connection with the course "Theory 1" at the Eastman School. Without a complete report, however, we cannot evaluate it...
> Larson makes the claim that the relationship of the tests to musicality is about as close as that between intelligence scores and school achievement. The statement has been made before, but the grounds for it are not known to me. Mean correlations between intelligence scores and marks run between .50 and .60, as Larson quite truly says. But the weighted average of the correlations I tabulated for the Seashore Tests
> with musical behavior (omitting my own low figures, since he objects to them, though I cannot accept his criticism), is but .35- quite a different story."

> "(2) If the tests do not measure general musical talent, what do they measure?
> ...No one familiar with the tests will doubt that, within the limits of their reliability, they do measure certain "hearing" abilities. But a very crucial question remains. Do these abilities operate in the same way in music as in the tests: Is pitch discrimination on the same basis when we are comparing two tones in isolation as when intervals and trends of tonality and melody are present
> ...One of their recognized features is that they avoid using musical materials. But may it not be, in consequence, that the mental processes they reveal are not those which occur in music? If not, they measure hearing, but not musical hearing. ... It is this point which must ultimately determine the practical usefulness of the tests."

This point probably best illustrates the omnibus

group's

main objection to the tests.

"(3) Are the tests practically useful? ...I need not remark that even though we show that a guidance program including the tests gives acceptable results, this in itself is no proof that they are valid for general talent, nor need it even tell us what they measure."

"(4) How should the tests be used? ... Clearly they should never be used as though they could immediately and certainly reveal a person's level of general or "omnibus" musical talent. ...Seashore contended that when they are employed, averages of the six measures should not be used, but rather a profile technique, in which each measure "stands on its own feet" as it were. However, there is a serious objection to this procedure. We give the tests to an individual, and find that he shows certain differences in respect to the six measures. But unless the tests are highly reliable, these obtained differences may be fictitious. ...Hence, there seems reason to question the profile technique, and to urge that the tests should not be used for individual diagnosis, but at the most for general "dragnet" purposes- a possible use mentioned by Seashore. And as is true of so many tests, they probably identify marked deficiency better and more surely than marked excellence."

"(5) Is there such a thing as general musical talent?
Seashore believes not, and has so stated in his article and elsewhere. He holds that musicality is not one single factor in the human mental make-up, but consists of a large number of specific and limited traits, of which the tests measure six. This is one representative view of the nature of human abilities. But the reader should know that it is far from being universal among competent psychologists. For myself I am unable to accept it. While it is clear that we must not think of musical talent as a sort of faculty, yet one may not unreasonably believe that all musical people,- pianists, violinists, timpanists, composers, keen listeners- have something in common. This we would call their "musicality" or their talent, and it might well consist of certain

> excellencies of hearing, innate and acquired-(perceptual configurations of unusual excellence)-which the Seashore Tests, dealing on the whole with sensory abilities, are not even designed to measure.
>
> All this has a very practical bearing. If there is no such unitary factor as musicality or musical ability, there is no use even trying to construct tests to measure it. This is Seashore's view. If the contrary is the case, then the way is hopefully open for research on the construction of tests different in principle from the Seashore tests. Such efforts are being widely made. What we have are really two working hypotheses for the direction of constructive research."

This last question contains within it the best description (that I have been able to locate) of the two different sides of the controversy, the Atomist and the Gestalt. That is one of the reasons why I consider these articles to be historically important. This description of the articles published in the Music Educators Journal has served to give the reader a sense of what the Atomist-Gestalt Controversy was all about. It's importance for the development of music aptitude research was in delineating what the crucial issues about music aptitude were that needed to be focused on. In the years that followed, researchers were able to take what seemed promising in terms of ideas from this discussion and leave behind the pieces that didn't seem to work as well.

Influences on Music Perception Research

Among the influential writings cited in research articles of the 60's,70's and 80's are Piaget's the Psychology of Intelligence(1950), which had a great influence on developmental psychology as a whole[29], Revesz's Introduction to the Psychology of Music (1954), Meyer's Emotion and Meaning in Music (1958), Chomsky's Aspects of the Theory of Syntax (1965),Goodman's Languages of Art (1968) Pflederer and Sechrest's Conservation-Type Responses of Children to Musical Stimuli,(1968), Shuter-Dyson's The Psychology of Musical Ability (1968), Attneave and Olson's Pitch as a medium: A new approach to psychophysical scaling (1971), Deutsch's Octave generalization and tune recognition (1972) and Harvard's Project Zero report on Basic Abilities Required for Understanding and Creation in the Arts (1972).[30]

---

[29]. Hargreaves,D.J. The Developmental Psychology of Music, Cambridge University Press, Cambridge, 1986

[30]. Revesz,G. Introduction to the Psychology of Music, Norman Oklahoma: University of Oklahoma Press. 1954

Meyer,L.B. Emotion and Meaning in Music, Chicago: University of Chicago Press, 1956

Chomsky,N., Aspects of the Theory of Syntax, MIT Press, Cambridge, MA 1965

Goodman,N. Languages of Art, Hackett, Indianapolis. 1976

Pflederer,M. & Sechrest,L. Conservation Type Responses of Children to Musical Stimuli,Council for Research in Music Education Bulletin, 13, 19-36, 1968

Shuter-Dyson,R. & Gabriel, C.,The Psychology of Musical Ability, (2nd. Ed.) London:Methuen 1981

Attneave,F. & Olson,R.K. Pitch as a Medium: a new approach to psychophysical scaling,American Journal of

Out of these articles, papers, books and others came an explosion in research in music cognition in the 70's and 80's. Harvard's Project Zero (1972) published a view of music perception as "an active organizing process in which the listener discovers or constructs his own coherence by spontaneously and/or deliberately coding the stimulus (not necessarily verbally) through various implicit means which can be described as categorical systems or languages."[31] In 1982, Dowling described music perception as a developmental sequence achieved through "melodic information processing"[32] and very recently, Sloboda adds the perspective that "The reality is that all music must reflect the psychological propensities and capacities of humans as composers, performers, and listeners." Saying in essence that we cannot isolate the learning experience from the "generative process" of creating or performing music.[33]

Psychology, 84,147-66. 1971

Deutsch,D. Octave generalization and tune recognition, Perception and Psychophysics, 11,411-412. 1972

31. Goodman, Perkins, Gardner, 1972 op.cit.p.24

32. Dowling, W.J. Melodic Information Processing and its Development, The Psychology of Music, Deutsch,D. Ed. Academic Press, 1982 p.413

33. Sloboda, J.A. Ed.,Generative Processes in Music, Clarendon Press, Oxford, 1988, vi

FOLLOW UP STUDIES

Seashore Measures of Musical Talents:

From the very beginning criticisms were leveled at the Seashore tests. The tests were originally presented on a 78rpm record that was played on a hand crank machine. To give a clearer picture of the problems that arose out that means of delivering the tests, I would like to quote two sources, first, informal comments made by Edwin Gordon, holder of the Seashore Chair in Music Education at Temple University, in a private interview and then an excerpt from the follow-up study done by Jensen and Gilliland in 1922:

> "... See, it's now on 33& 1/3 rpm. The last version, 1956 by Psychological Corporation is as 33 & 1/3. The reason being that most mechanical record players, which most of them were when Seashore started in 1919 were wind up. Now when you go to that they will vary 10-12%....in other words what they're saying is because turntable speeds are inconsistent, the test results are invalid... that's why Seashore went (away from the hand crank machines) in 1939. When the electric phonographs came out, there was some improvement there but there was still a difficulty and finally in 1956 he went to the 33 & 1/3 rpm with Psychological Corporation and he had hoped at that time that it would be more consistent but I can tell you even with the 33 & 1/3 rpm, it's not good. And that's what has invalidated the Seashore tests, one of the things."[1]

Back in 1922 the focus of the complaints was slightly different. Jensen and Gilliland seemed to accept the phonograph as a piece of modern technology that was

---

[1] Gordon, E. 1986 Private Interview. Philadelphia, PA

suitable for classroom use but too scratchy for use in a laboratory experiment. The issue of the turntable speed is not mentioned at all:

> "While giving the Seashore tests for pitch, intensity, time, consonance and for tonal memory as arranged for the Columbia graphophone records certain distracting factors seem to vitiate the results. One of these distracting factors was the scratch of the needle on the phonograph record, another was the always more or less noticeable metallic overtones common to all tones reproduced on the phonograph. It seems reasonable to think that these two factors might seriously affect the student's judgement of pitch."[2]

The criticisms of the test, its means of application and its methods of ascertaining reliability and validity became more centered on the test construction and conception as time went on.

James L. Mursell, one of the editors of the Silver, Burdett series on music education and a professor of education at Columbia Teachers College did a follow up study of Seashore's test. Mursell's study, which he reported in two articles, came out in 1931/32, long before the controversy heated up in the pages of the Music Educators Journal. In the second article he published further results of the follow-up study: "Measuring Musical Ability and Achievement: A study of the correlations of Seashore test scores and other

---

[2]. The Reliability of the Seashore Phonograph Record for the Measurement of Pitch Discrimination, Jensen, C.R. & Gilliland, A.R., Journal of Experimental Psychology, 1922, Vol. 5

variables."[3] I include it here because it formed the basis of a dialogue between Mursell and Seashore and on a larger basis, between the Atomist group and the Gestalt group.

Mursell's study was undertaken in order to establish what were appropriate criteria for validation of tests of music aptitude and in order to answer the question: do the Seashore tests really measure musical ability?. In order to do this, he gave the test to 176 students at Lawrence College, 88 in the music program and 88 from the general student body. He included in the overall rating of the music students a percentage of their grade in piano or voice and ratings for talent by every instructor of "applied music" they had studied with for at least one semester. He calculated intercorrelations for his figures along with correlations from studies done by Ruch and Stoddard, Brown and Seashore[4] and then applied the Spearman-Brown

---

[3]. Mursell, J.L. Measuring Musical Ability and Achievement: A study of the correlations of Seashore test scores and other variables, Journal of Educational Research, Vol. 25 No. 2, Feb. 1932.

See also: Mursell, J.L., An Evaluation of the Seashore Tests, Music Supervisors Journal, May, 1931, p.62-65

[4]. Ruch, G.M. and Stoddard, G.D. Tests and Measurements in High School Instruction. Yonkers, New York: World Book Co. 1928

Brown, A.W. The Reliability and Validity of the Seashore Tests of Musical Talent, Journal of Applied Psychology, XII (1928) pp.468-476.

prophesy formula to them. He defended this procedure in the following manner:

> "The applicability of this formula to the Seashore Tests has been debated by Farnsworth and Lanier. There seems to be no serious objection to its use. Indeed, it may even be argued that the correlation of comparable halves will yield a better measure of reliability than self-correlation in toto, especially if the tests are repeated only a short time after the first administration. This is due to their fatiguing and monotonous character."[5]

On the basis of these combined scores he went on to compute reliability coefficients for the Seashore tests and came to the following conclusions:

> "1. The Seashore Tests were found to have low intercorrelations, confirming previous studies. This may indicate either that there is no such thing as unitary musical capacity, or that these tests do not measure it.
>
> 2. Reliabilities were found by correlating the first and second five rows of the scoring tables and applying the Spearman-Brown prophecy formula. This appears defensible, and may be even better than self correlation of the complete tests.
>
> 3. Reliabilities so obtained were approximately similar to those of previous studies, thus confirming the testing procedure.
>
> 4. Reliabilities as revealed by the standard error were so low that the use of the tests for individual diagnosis seemed questionable.
>
> 5. In most of the tests, the performance of conservatory students was superior to that of college students, measured both by averages or median overlapping. This superiority was not sufficiently marked to warrant any very specific educational advice.
>
> 6. Grades in Applied Music, and estimates of musical talent seemed to have a reliability which

---

Seashore, C.E. and Mount, G.H. Correlation Factors in Musical Talent and Training, Psychological Monographs, XXV (1918) pp. 47-92.

23. Mursell,1932 op.cit.

should render them good validation material.

7. There was almost no relationship between Seashore Test performance, and the above criteria. Taken with similar results from other studies, this leads to the conclusion that the tests cannot make fine discriminations of true musical talent within musical groups.

8. The relationship of the Seashore Test scores to performance on special tests indicates that the former may be of use as aids in diagnosing special musical abilities.

9. Where significant correlations between Seashore Test scores and tests of special musical abilities and types of achievement are found, we are usually dealing with very heterogeneous groups, which indicates that the battery may be able to discriminate roughly, though it cannot do so very accurately or finely."

Hazel Stanton was a student of Seashore's who carried out some follow-up studies of Seashore's tests at the Eastman School of Music under the auspices of William Larson, head of the music department, and Seashore.[6] The studies lasted for ten years and included: tests for children & adults; administering both the Seashore tests as well as an IQ test; The Iowa Comprehension Test, which is described as a group test of general intelligence. Mursell described the Stanton studies in the following way:

"However, we have not yet considered the most

---

[6]. See Stanton, Hazel M., The measurement of musical talent: the Eastman experiment. University of Iowa Studies in the Psychology of Music, Vol.2, Iowa City: University of Iowa Press, 1935

Stanton, H., Koerth, W. Musical Capacity Measures of Adults Repeated after Musical Education. University of Iowa Studies. Series on Aims and Progress of Research, 1930, No. 31

important and ambitious validation study on the Seashore tests, the ten year experiment in their use carried on at the Eastman School of Music by Hazel Stanton. ...On the basis of this battery individuals were segregated into five classes- "discouraged, doubtful, possible, probable, and safe," and their later achievement in the conservatory was studied. Annual academic survival, avoidance of dismissal, attainment of scholarships and honors, recital appearances, and graduation were the most important factors considered. In all these respects an increasing degree of success was demonstrated in passing from the low to the high groupings."[7]

Mursell went on to question the analysis of the results and also the methods used to obtain a final rating on the battery of tests as designed by Stanton. In brief, talent profiles were made up from the results of the Seashore tests then the scores from each test were given a letter grade. Then each profile as a whole was given a letter grade on the basis of the five grades it already contained which yielded a classification system of talents. Teachers ratings for individual pupils were included in the tests and letter ratings were given to the intelligence test scores. Then the whole system was classified as either discouraged, doubtful, possible, probable, or safe. Stanton called this classification system a "cumulative key" and described it this way:

"From accumulated data and from observation, it was noted that certain students could be depended upon to achieve the four-year program creditably and with satisfaction to themselves and to the faculty of

---

[7]. Mursell, J.L., The Psychology of Music, Greenwood Press, Westport, CT. 1971

the music school. These students could be considered a safe academic risk, hence it seemed natural to designate them as a "safe" group. Another group had less of a margin of safety and yet under certain conditions they probably would succeed in making satisfactory musical progress. These students naturally fell into a "probable" group. A third group of students would find musical progress possible but the odds against them would be greater, hence this group was designated "possible". A fourth group consisted of those who were doubtful risks; those who, for various reasons, would not, with few exceptions, carry the work of the course with sufficient credit or satisfaction to warrant the effort involved. "Doubtful" was the natural term to designate this group. A final group consisted of those who were obviously not fitted to carry on regular course work in a music school. The odds against them in this particular field were too great to justify encouragement. This group, therefore, was called "discouraged". Hence a five-fold classification of Safe, Probable, Possible, Doubtful, and Discouraged for entrance into the regular course of four years leading to a bachelor of music degree was devised as a useful administrative tool for the selection and classification of students."[8]

This classification system used in conjunction with the Seashore tests were the basis of Stanton's follow up study and her conclusions about the reliability and validity of the Seashore tests. According to Mursell,

"It is difficult to form a reliable opinion of the statistical meaning of this rather involved and decidedly unusual procedure. But one thing is abundantly clear. We have here nothing in the way of a direct validation of the Seashore Measures of Musical Talent. The results are evidently of considerable practical value. But they are based on factors elaborately combined and nowhere analyzed in isolation from each other. They furnish no proof whatever that the Seashore tests given independently of any other measures will yield a valid index of

---

[8]. Stanton, H. 1935, op. cit. p.69

musical capacity."[9]

Why Mursell says that Stanton's follow-up study is of "considerable practical value" is not clear. My guess would be that he is trying to separate his criticisms of Seashore from being considered to harsh to Hazel Stanton personally. Based on her ten years of research using the Seashore Measures, Stanton concludes that the Measures are valuable in assessing talent in students and predicting their success in a music school. She cites the following reasons why the Seashore tests should be used:

> "The Measures aid (1) in the placement of musical instruments with the more talented children, (2) in recommendations to parents concerning the purchase of instruments, (3) in segregation of instrumental classes for more homogeneous groupings of differential talent, (4) in cooperation with music teachers in the study of unusual cases, (5) as a check on accomplishment in music classes in an effort to have progress conform with aptitudes, (6) in cooperation with various organizations interested in child guidance, (7) in the limitation of instrumental classes to those who can profit by the instruction."[10]

Despite the practicality of Stanton's intentions, there were other problems with the Stanton study. Holstrom (1963) points out that instead of the Seashore tests measuring musical capacities which were resistant

---

9. The Psychology of Music, Mursell,J.L., W.W.Norton & Co. New York. p.299

10. Stanton, H. 1935 op. cit. p. 138

to experience, the high level of pitch perception already present in the selected group of subjects Stanton used in her tests accounts for their lack of improvement over time.[11] Other critics have pointed out that the use of an IQ test in the Stanton study confused the results because it was included in the cumulative key. Their point being that the IQ test and an audition might have done just as well in predicting success without the Seashore Measures.[12]

Taylor did a 5-year follow-up study of the Seashore and Kwalwasser-Dykema tests at the College of Music in Cincinnati from 1930 to 1935. She used the two tests mentioned above plus an original test she devised for the experiment and an I.Q. test; the Detroit Advanced Intelligence Test. She included music course grades and teacher evaluations. After the experiment was over, in 1939, she did a survey of the graduated students who had participated in the study. She concluded that the predictive value of the Seashore tests was very low. "No coefficient is higher than .30." She grouped the survey data in 1939 into five areas: group I consisted of those who had become teachers of music at the college level, directors of music programs in schools or large radio stations,

---

11. Holstrom, L.G. 1963 op. cit. p.168

12. Lehman, 1968, op. cit. p.42

classical or symphony performers, conductors or composers and students pursuing advanced degrees in music; group II consisted of elementary or secondary school music teachers who were doing well, private teachers of music, radio performers, director of music in a large church and students who possessed a BA in music plus graduate study; group III consisted of mediocre teachers of music in schools, private music teachers of pupils of average attainment, performers in "spot" radio shows and occasional casual engagements and paid soloists in a church choir; group IV was made up of teachers and performers of music who were just barely making it in the field and; group V was persons who had changed to other professions due to lack of success in music and students who had failed music in college.[13]

Taylor concluded from this study that:

> "It is evident that, as a whole, the music test batteries do not evidence sufficient predictive power to be used by themselves for guidance purposes, yet neither do they have so little value as to warrant discarding them entirely."[14]

She added that the students who did best in sight singing and dictation courses in college did the best as professionals in music.

---

[13]. Taylor, E.M. A Study in the Prognosis of Musical Talent, Journal of Experimental Education, 1941

[14]. Taylor, E.M. 1941 op. cit. p.27

Manor (1950) did a study combining the Seashore tests of pitch, tonal memory and rhythm with an intelligence test, a teacher evaluation and a training period of eight weeks. After that the subjects took a fourteen week course in an instrument of their choice. Manor found low correlations (.21-.49) between the Seashore tests and success at instrumental or vocal music:

> "Of the three Seashore tests, the Pitch Test showed the closer relation to the achievement scores, a correlation of .49, also the highest correlation found between the guidance measures and achievement.
> ...Although the measures might be helpful in a guidance program, rigid differential individual guidance cannot be justified on the basis of the results found in this study."[15]

I think that Manor is referring to Seashores "Talent Profiles"[16] when writing about "rigid differential individual guidance". I also found it interesting that by 1950 Manor as well as other researchers in music aptitude had become a little more cautious in their claims regarding music aptitude tests; Manor opens the report of his study by saying "Musical talent, because of its complexity, probably cannot be measured by any

---

15. A Study in Prognosis: The Guidance Value of Selected Measures of Musical Aptitude, Intelligence, Persistence, and Achievement in Tonette and Adaption Classes for Prospective Instrumental Students. Manor, H.C., The Journal of Educational Psychology, 1950, p.42 & 44.

16. See p.316 in Psychology of Music, Seashore, C.E. 1938

one test or scale."[17]

Kyme (1956) opened his follow-up study report by saying:

> "Musical performance is a complex process which involves the integration of the elements of music-pitch, timbre, loudness, and duration- into meaningful wholes. It occurred to this investigator that a test of musical capacity which could somehow get at this ability to integrate might be superior to tests which involve only the perception of differences in the elements of music taken in isolation."[18]

He devised a test of aesthetic judgement which was given to twenty-seven judges of performance by marching bands playing at a music festival. The performances were recorded, the judges heard three different pairs of performances of one minute duration. The judge was then to indicate whether the versions were the same or different, his preference and an aesthetic ranking of the performances. The correlations between the judges scores and teacher rankings were high, (.58-.82). He then compared his coefficients for his aesthetic test to the Seashore, K-D and Whistler-Thorpe tests and found that his correlations were much higher. For instance, .08-.46 for Seashore's tests as opposed to .74 for his. He made the following conclusions about his test and testing in music aptitude in general:

---

17. Manor, H.C. 1950 op. cit. p.31

18. Are Musical Tastes Indicative of Musical Capacity? Journal of Research in Music Education, Kyme, G.H., 4, 44-51, 1956, p.44

> "It (his test) has established that aesthetic judgments, functioning as organizing factors of auditory images, may be used to differentiate between persons known to be musical and persons observed to be less musical. ...Finally, it has marshalled some evidence that the best prediction of musical behavior must be based upon observations of the subject in. many musical situations, and that the act of appreciation, that is, the assessment of music at its true value by the individual in the light of his experience, is worthy of more consideration than has been given it heretofore."[19]

In reviewing these follow-up studies I have tried to give a sense of how the thinking of researchers in the field was changing over the years from 1930 to 1956. The trends of the larger world of psychology can be seen in the perspectives and interests of the individual studies. The Seashore tests, definitely the most used and written about tests of musical aptitude, were seen as too limited in scope as time went on and Gestalt influenced theories came more into prominence.

Kwalwasser-Dykema Music Tests:

The K-D tests were found to have low reliability coefficients by Taylor (.20-.61) Farnsworth (.40), Whitley (.22-.83)and Beinstock.[20] To summarize the

---

19. Kyme,G.H., 1956, op. cit. p.51

20. Taylor, E.M. 1941 op. cit.1-28
Farnsworth,P.R. Studies in the Psychology of Tone, Genetic Psychological Monographs, 1934, 15:1, 1-91
Whitley, M.T., A comparison of the Seashore and K-D. tests. Teachers College Record, 1932, 8, 731-51
Beinstock, S.F. A predictive Study of

findings of Bienstock: She found low reliabilities for the K-D tests when administered to 79 high school aged children. She found that although the teachers were in agreement about the music and academic scores of the students, the test-retest correlations of the K-D tests after a 12-month interval of musical training were very low with the one exception of the tonal memory subtest. It had a reliability coefficient of .52; The only test in the K-D battery to have statistically significant reliabilities. She put forth the possibility that the reliabilities were low because the population of the test was highly selected for their success in music. She came to the following conclusions at the end of her report:

> "1. The Kwalwasser-Dykema tests were too unreliable to be used for the prediction of individual success in music. Improvement in the reliability of the Rhythm and the Tonal Memory tests might make these tests practicable for guidance purposes.
>
> 2. The interval of one year of musical training in the Music and Art High School did not reliably increase the scores on the K-D battery.
>
> 3. There was a positive correlational trend between the Kwalwasser-Dykema tests and success in theoretic and applied music but it was too low to be of practical value.
>
> 4. Teachers' marks used as criteria in this study had adequate reliability for use as measures of achievement.
>
> 5. The individual performance test used in this

Musical Achievement, Journal of Genetic Psychology, 1942, 61, 136-45 study yielded only slightly higher coefficients than the K-D tests when correlated with the music marks, and was unsatisfactory for prognosticating musical achievement.

6. Marks in applied music (musical performance) were generally prognosticated with much less accuracy than were marks in theoretic music.

7. The most effective measures for the prediction of success in theoretic music were the intelligence quotient and the age of the students, while the least contributive was the extent of prior music training and the performance test score.

8. The instrumental students were, as a group, superior to the vocal students on all the predictive measures, and also on all directly comparable measures of achievement.

9. The findings from the analysis of the records of the highest and the lowest ranking students conformed in general to the trends observed for the group as a whole. The measures were more effective for predicting failure than for predicting success. As a rule very low scores on several of the predictive measures were almost certain to indicate poor achievement, whereas high initial scores did not insure superior achievement."[21]

Farnsworth found the K-D tests too unreliable for individual prediction of success in music except for the tonal memory subtest.[22] In 1954 Holmes published a revised version of the Kwalwasser-Dykema tests. In it he reported higher reliabilities for the tests than had previously been found by other researchers. He based his results on making "systematic and fundamental

---

[21]. A Predictive Study of Musical Achievement, Journal of Genetic Psychology, Bienstick, S.F., 1942, 61, 135-145.

[22]. Studies in the Psychology of Tone, Genetic Psychological Monographs, Farnsworth, P.R. 1934, 15:1, 1-91.

changes in the administrative directions and scoring keys" of the tests.[23] His criticisms of the tests were as follows:

> "...the scale consists of 25 pairs of patterns which vary in length from four to nine tones. The subject judges whether the pair consists of the same or different halves. Discrimination is therefore not only very rough, but the subject is "structured" to guess-and, by the dichotomous nature of the possibilities, half of his guesses can be correct. Further, since the manual says, "Plenty of opportunity should be afforded the student to hear the first few items of the different tests and to ask questions...," and since each of the tests contains some glaringly obvious items, the combined effect is to reduce the number of items working to discriminate to something below 12. These conditions are common to all of the scales."[24]

Unfortunately, for all the work he put into redesigning the tests, changing the format of the responses from true-false to multiple choice, thereby making it possible for the subject to make much finer discriminations in answering, still the tests were reported as having reliabilities too low to be used for prognosis of success in music.[25] With reliabilities being as low as they were (only 25 pairs of melodies) than the validity of the tests comes into question.

The Drake Musical Aptitude Tests:

---

23. Increased Reliabilities, New Keys, and Norms for a Modified Kwalwasser-Dykema Test of Musical Aptitudes. The Journal of Genetic Psychology, Holmes, J.A., 1954, 85, 65-73

24. Holmes, J.A., 1954, op. cit. p.66

25. Lehman, 1968, op. cit. p.46

The Drake Music Tests consisted originally of four tests: musical memory, interval discrimination, retentivity and intuition. Of the four, only musical memory and interval discrimination were found by Drake to be useful (Drake, 1933).

He made a recording of the musical memory test in 1942 and a rhythm test in 1954. Lundin (1949) found the validity figures to be lower than Drake had predicted. Gordon studied the effects on training and practice (1961). He tested 65 students of high school and college age non-music students. Out of these he chose twenty to be tested and trained further. Of the twenty subjects, ten had scored high on the Drake test and ten had scored low. He divided these two groups into groups of 5 high and 5 low. One group of ten high/low subjects became the control group. The other ten high/low subjects became the experimental group. The experimental group received 20 training periods for a half of an hour each. Gordon devised original musical phrases similar to the Drake tests and the subjects were then taught how to listen for changes. The last three trainings were used for training in rhythm discrimination. The training lasted for one month. Then both the control group, who had not received training, and the experimental group took the Drake tests again. Gordon summarized his conclusions this way:

> "Within the limits of the precision of this experiment no conclusive effect of practice was revealed. Thus, tentative support for Drake's assertion of the test's insensitivity to training may be derived from these results. However, the obtained difference, while not significant, was consistent with the hypothesized effect of training. Also, the informal tests built and used by the instructor during the training period gave some indication of growth in skill. It is possible that increased experimental precision -gained by the use of larger groups- might reveal similar size observed differences to be statistically significant."[25]

Ferrell did a follow up study in which he found that the test successfully discriminated between students with a high level of musical aptitude and those without.[26] Most researchers reporting on the Drake tests gave it slightly more favorable reviews than the Seashore and K-D tests.[27]

Wing Standardized Tests of Musical Intelligence:

The Wing Standardized Tests of Musical Intelligence (1948), consisted of seven tests, pitch, memory,

---

25. A Study to Determine the Effects of Training and Practice on Drake Musical Aptitude Test Scores, Journal of Research in Music Education, Gordon, E.,9 (I), 63-74, 1961

26. Ferrell, J.W., A Validity Investigation of the Drake Musical Aptitude Tests, PHD Thesis, Music, State University of Iowa, 1961

27. Lundin, 1953 op. cit. P. 255

The Social Psychology of Music, Farnsworth, P.R. Iowa State University Press, 1969 p.202

Lehman, 1968 op. cit. p.52

rhythm, harmony,chord analysis, intensity and phrasing. The chord analysis test consisted of 20 items in which the subject was asked to count how many notes in the chord. In the other tests subjects were asked to make 'same-different' comparisons from pairs of examples.

Bentley (1955) correlated the Wing tests with four other music aptitude tests and an IQ test. He found the Wing tests to have high reliabilities (.857) particularly for the Melodic Memory test and secondarily for the Pitch Change test. He found the recordings though, to be inferior to other test recordings. He concludes:

> "Three test authors, Gaston, Whistler-Thorpe, and Wing measure an interaction of acuity and environmental acquired ability. These three tests (Gaston, A Test of Musicality, Whistler-Thorpe, Musical Aptitude Test, and Wing, Tests of Musical Ability and Appreciation) are constructed as measures of various aspects of musicality."[28]

A comparative study was done by Heller (1962) to study the effects of formal music training on Wing test scores. More studies comparing the Wing tests to other tests were done in the 1960's, Cain (1960) did a comparison study of the Wing test to the Gaston and Drake tests. Heim (1963) studied the musical aptitude of blind subjects using the Wing test and Wertz (1963)

---

28. Bentley, R.R., A Critical Comparison of Certain Music Aptitude Tests, EdD thesis, University of Southern California, 1955, p.311

studied the relation between changes in musical preference and scores on the Wing test.[29] Lehman sums up the prevailing views by music researchers on the Wing test in this way:

> "The battery is based on an omnibus approach, reflecting Wing's belief that there exists a general factor of musical talent. The tests have been carefully prepared, with the result that the Wing battery represents one of the better instruments available today for the measurement of musical ability. Perhaps its most serious limitation is the technical quality of the tape. Also, when the test items are supposed to be the same, it would be better to re-use the original master recording of the excerpt than to ask the performer to play the excerpt again in an identical manner. The battery deserves to be re-recorded in order that it can be of maximum benefit to music instructors and psychologists. The reliability and validity should be investigated further, and percentile norms, though less frequently used in Britain than in this country, should be made

---

[29]. Bentley,R.R., 1955, op. cit.

Heller, J.J., The Effects of Formal Music Training on the Wing Musical Intelligence Scores. Doctor's thesis, State University of Iowa, (Iowa City, Iowa) 1962 (DA 23:2936)

Cain,M.L., A Comparison of the Wing Standardized Tests of Musical Intelligence with a Test of Musicality by Gaston and the Drake Musical Aptitude Tests. Master's thesis, University of Kansas (Lawrence, Kansas) 1960.

Heim,K.E., Musical Aptitude of Senior High School Students in Residential Schools for the Blind as Measured by the Wing Standardized Tests of Musical Intelligence. Master's thesis, University of Kansas (Lawrence, Kansas), 1963.

Wertz, C.B.,Relation of Changes in Musical Preference to Scores on the Wing Standardized Tests of Musical Ability and Appreciation. Master's thesis, University of Kansas, (Lawrence, Kansas). 1963 available."[30]

All in all, of the four tests which I have presented follow-up studies for, the main factors which are criticized are: the reliabilities; the validity;[31] the design of the responses and administration,(in the case of the K-D tests); and in the case of Seashore in particular, the limitations of design of the stimulus (pairs of tones instead of melodies). What follows on the next page is a table illustrating the main criticisms of the four tests. I have included reliability and validity figures from follow-up studies which I did not cover in this paper because I wanted to give as much of a picture of the range of figures as I could.

For the most part, reliability and validity is low for most of the tests but definitely for the Seashore and K-D tests. It should be noted though, that many more studies were done with the Seashore and K-D tests than with the Drake and Wing tests. As a result, the figures can be misleading. If there were more follow-

---

30. Lehman, 1968 op. cit. p.48

31. Standards for what constituted validity varied from test to test but centered for the most part around issues of size of population tested, number of items that make up a subtest, technical aspects like quality of recordings, and whether the design of a particular stimulus actually measured what it claimed to measure.

Tester:	Reliability	Validity	Technology	Stimulus	Response	Theoretical B.G.
Seashores Measures of Musical Talents	.08 (Mursell) .30 (Taylor) .26 (Drake) .21-49 (Manor) .08-.46 (Kyme)	.26 (Drake) .08 (Mursell) .35 (Brown) .42 (Taylor) .46 (Kyme)	Turntable speed varies. Records are too scratchy.	Test bear no relation to music behavior. Pairs of tones too limited. Meas. acoustic Properties only		Atomist
Kwalwasser-Dykema Music Tests:	.09-.39(Bienstoc .20-.61 (Taylor) .40 or less (Farnsworth) .22-.83 (Whitley)	.01 (Chadwick) .23 (Drake) -.23 (Tilson) -.06 (Taylor)		Not structured to measure fine discrimination.	Redesigned from T/F to multiple choice.	Atomist/w some Gestalt
Drake Music Aptitude Tests:		.58 (Lundin) .55 (Farnsworth) .55 (Lehman)				Omnibus/Gestalt
Wing Standardized Tests of Music Aptitude:	.65-.86 (McLeish)		Poor quality of record & tape.	Reliability of repeated live performance of stimuli in doubt.		Omnibus/Gestalt

up studies done with the Drake and Wing tests, there would probably be some that showed low reliabilities. Also, the criticisms of the design of the tests were primarily focused on the Seashore and K-D tests. One reason for this might have been that they exemplified the atomistic design of tests and became obvious targets of researchers wanting to try a Gestalt approach to testing. Another reason for the majority of follow-up studies done with the Seashore and K-D tests is that they were the most widely used tests in the U.S. In the following chapter, Comparison of the Pitch Subtests, I am going to compare the pitch subtests with more recent experiments in pitch perception and melodic processing.

COMPARISON OF THE PITCH SUBTESTS

In this chapter I am going to take the following subtests from the four tests discussed in the proceeding chapters and compare them to newer research in pitch perception and melodic processing to see what has changed and what has been learned about pitch perception in the intervening years: Seashore's pitch and tonal memory tests; Kwalwasser's pitch and tonal memory tests; Drake's interval discrimination, retentivity, and musical memory tests; and Wing's pitch, chord analysis, memory and harmony tests. I have picked these subtests because I think that most of the subsequent research done in the field has used one of these forms of musical stimuli. I am interested in seeing which form of stimuli is the most widely used.

The changes in the field of psychology which came to be known as the new "cognitive science" were still just becoming known in the 1960's. During the mid to late 1950's and early 1960's there weren't a lot of experiments being done in music perception or aptitude aside from the follow-up studies discussed above which seemed to be running out of steam. It took some younger psychologists familiar with the newer theories coming out of cognitive psychology, particularly neurophysiological research, linguistics, information processing and the publishing of Neisser's Cognitive

Psychology[1] to start generating some new ways of looking at music perception.

One of the first of these new ways of looking at music perception was a study done by Pflederer and Sechrest (1968). Basing their research methodology on Piaget's work, they set about to study conservation of music in the cognitive development of children. This was an approach to research in music perception which had been largely ignored up until that point. In the chapter on Alternative Methods of Assessing Musical Aptitude, I describe this experiment in much more detail, but for the sake of pitch comparisons, I will describe it here very briefly. 198 children were randomly assigned to experimental and control groups. They listened to musical phrases which had been altered in one way or another and they were to describe to the experimenter whether the musical phrases were the same or different. The method for recording the data was like that of a Piagetian style protocol; the examiner asks questions and the subjects response is notated. What they learned from their research was that children use a variety of terms to describe music. These terms can be categorized based on the following: (A) Method and/or medium of performance; (B) simple description

---

[1]. Neisser, U. Cognitive Psychology, Appleton, Century, Crofts, N.Y. 1967

(often, but not always, accurate) of the musical stimulus; (C) Affective reaction; (D) Imagery, including responses suggestive of synesthesia, (E) Reference to rhythmic movement."[2] It was found that children sometimes conclude that a change in tempo, mode or rhythm of a melody constitutes an error in performance rather than as an alternative way of playing the tune. Also, the addition of harmony was sometimes seen as a change in timbre and sometimes as a change in tempo. Lower scale degrees were perceived as being closer in pitch and performed faster. Changes in tempo were also seen by the children as being a change in the overall quantity of the music or that the length of the piece had been increased. The responses of the children in this study were found to be like the Piagetian speed experiments in that evidence of preoperational thought occurred in children up to thirteen years of age. This happened particularly when perceiving changes in tempo as changes in the overall quantity of music played. This study gave rise to new methods of experimentation in the field of music perception.

Attneave and Olson (1971) presented researchers with other new perspectives in their study of

---

[2]. Pflederer, M. & Sechrest, L. Responses of Children to Musical Stimuli, <u>Council for Research in Music Education Bulletin</u>, 13, 19-36, 1968

psychophysical scaling.[3] Their comments about the field of psychophysics are telling in terms of the kinds of shift in perspective that were going on:

> "Psychophysics has a curiously autonomous status in contemporary psychology. The measurement of elementary sensations is obviously important if sensations are viewed as the building blocks of perception. Since this view has not been widely entertained for some five decades, however, one is entitled to wonder why psychophysicists go on measuring sensations with undiminished zeal. The basic fact that perception is relationally determined- that perceptual objects owe their identity to certain relational invariants of patterns or configurations- has profound consequences for the psychophysical treatment of such commonly investigated continua as brightness, loudness, and pitch. A major defect of conventional psychophysics is its failure to take these consequences into account."[4]

Attneave goes on to describe pitch not as a perceptual object, the conventional way that psychophysicists see pitch, but as a pattern which is transposable, "a medium in which the same pattern may have different locations."[5] The subjects in the experiment heard two tones which were manipulated into random patterns of twelve or thirteen notes that stayed within an octave. The subjects task was to match the pattern heard in a different frequency region. All of the tones were played on an oscillator. This was done

---

3. Attneave, F. & Olson, R.K., Pitch as a medium: A new approach to psychophysical scaling. American Journal of Psychology, 1971, 84, 147-166

4. Attneave & Olson, 1971 op. cit. p.147

5. Attneave & Olson, 1971 op. cit. p.148

to control any affects of overtones. The outcome of the experiment was that non-musical subjects showed an increase in errors in the low frequencies and the musical subjects showed an increase in errors in the high frequencies. They conclude:

> "These data strongly support the view that intervals may be represented or encoded in either an analog or a digital manner. The differential effect of interval size on variability suggests that the nonmusical subjects were representing intervals as quantities or magnitudes or distances (on what scale is an open question), whereas the musical subjects were representing them in qualitative or categorical terms up to about 5000 Hz, but quantitatively at higher frequencies. Musical categories thus may be conceived as rigidly interconnected with one another in something like a solfeggio reference structure, the whole of which is freely transposable (below 5000 Hz) in a log frequency medium."[6]

The interesting thing here is that his results point out the perceptual differences reported between the musical and non-musical subjects. Also, they are presenting a new way of looking at subjects "internal representation" of musical structures. They postulated that the problems that the non-musical subjects had with the patterns was based on the lack of musicality of the two note pattern. They also thought that the musical subjects did better by "digital coding-by assigning tones to pigeonholes with determinate separations in the log frequency medium." They ended up designing a subsequent test using the NBC chimes as

---

6. Attneave & Olson, 1971 op. cit. p.158

their melodic pattern. The task was basically the same. On the second test the non-musical subjects did much better. Attneave & Olson concluded that non-musical subjects could be successful at the task as long as they were given "a well defined, highly overlearned pattern as standard."[7]

In 1968 Deutsch published a paper on music recognition in which she proposed some new theories based on neurophysiological research. Her theory was that music is processed through two parallel channels, "A & B":

> "In the first stage of transformation on channel A, primary neurons feed in twos and threes on the second-order neurons. These second-order neurons thus respond to specific intervals and chords. In the second stage of transformation second-order neurons are linked to third-order neurons in such a way that all units activated by seconds feed into one unit, all those activated by thirds into another, those activated by a particular triad into another, and so on. Thus the third-order neurons respond to abstracted intervals and chords."[8]

Deutsch was influenced in her theories by the British empiricist view of information processing done by Broadbent and Cherry (1953-1954) they were the first psychologists to represent cognitive functioning with a flow chart. Deutsch's research is characterized by her "bottom-up/molecular" (per Gardner 1985) approach to

---

[7]. Attneave & Olson, 1971 op. cit.

[8]. Deutsch, D. Music Recognition, Psychological Review, 1969, Vol. 76, No.3, 300-307

cognitive research. Her reports of experiments are widely read and she is considered one of the leaders in the field of music perception research.[9] All through the 1970's and 1980's she published results of her experiments.[10] In 1978 she published a report of an experiment called "Delayed pitch comparisons and the principle of proximity."[11] The principle of proximity refers to the Gestalt concept that objects close to each other but not necessarily related may appear as a group. In Deutsch's experiment she postulated that the principle of proximity applied to tonal sequences would mean that melodic sequences of smaller intervals would be more easily processed than sequences of larger intervals. She found support for this view in the Attneave & Olson paper just described, among others. Her stimuli consisted of two test tones which were to be compared by the subject to determine if the second tone was higher or lower than the first. Interpolated between the two tones was a sequence of tones which created a distraction to the discrimination of the test tones. The point was that if the interpolated

---

[9]. Gardner, 1985 op. cit. p.91

[10]. See bibliography for a complete listing of experiments reviewed.

[11]. Deutsch, D. Delayed pitch comparisons and the principle of proximity, _Perception & Psychophysics_, 1978, Vol. 23, (3) 227-230

sequences were made up of smaller intervals that they would be more easily processed and the subject would make fewer errors. Her results supported this hypothesis. In the chapter entitled Alternative Methods of Assessing Musical Aptitude, I have included a description of this report in detail. I will be using the data from that experiment in the comparisons of old and new forms of testing later in this chapter.

Another important paper for the field of music aptitude came out in 1970. Dowling and Fujitani had done two experiments investigating the role of "melodic contour recognition in memory for melodies". Their hypothesis was that:

> "The pattern of relationships among tones in a melody is what is important, and not their absolutely defined pitches. Hence, a given sequence of tones remains the same melody if each pitch is changed by the same amount. In musical terms, a melody is unchanged by transposition to a new key."[12]

They designed their tests to highlight the preservation of contour, (the set of directional relationships between successive tones in a melody) through changes in interval size and transposition. To do this they designed their test in the following way:

> "In Expt. 1 three groups of subjects were given different tasks. One group heard a standard melody (different for each trial) and after a 2-sec delay

---

[12]. Dowling,W.J. & Fujitani,D.S. Contour, Interval, and Pitch Recognition in Memory for Melodies,Journal of the Acoustical Society of America, 1971, Vol 49, No. 2 (2), 524-531

heard either an exactly identical comparison melody or a random collection of notes. The second group heard the standard and then either the same melody again, or a comparison melody with different notes and interval sizes but the same contour (ups and downs). these two groups were told to judge whether the comparison melody was identical to the standard or not. The third group heard the standard melody and then either a comparison with the same contour as the standard but different notes and intervals, or a random collection of notes. This third group was told to judge whether the comparison had the same contour as the standard."[13]

What they found was the following:

1. The easiest part of the tests was distinguishing same from random comparison melodies.

2. In the untransposed condition, it was easier to recognize same comparisons than same contour comparisons. They concluded that subjects were using recognition of pitches to solve this.

3. In the transposed condition, contour was the basis for recognition. The subjects ability to compare same contour with standard as opposed to standard from random was done with about equal proficiency.

4. Exact interval recognition was not used as a strategy by subjects to discriminate between melodies. This was proved by the fact that discrimination between transposed same and same contour comparisons was not good.

5. The precision of pitch and interval judgment

---

[13]. Dowling, W.J. & Fujitani, D.S., 1970 op.cit. p.526

encouraged by musical training is of little help in recognizing melodic contour.[14]

There is more detailed descriptions of the other aspects of this experiment and the 2nd experiment in their report. For the purposes of comparing stimulus design for pitch perception, this description I have given has all of the needed information.

Moving ahead ten years, another important report was published in 1980 by Massaro, Kallman & Kelly.[15] The purpose of the experiment was to see whether and how three "auditory characteristics" function in melody recognition. The characteristics were: tone height, a tone's frequency; tone chroma, position of a note within an octave; and melodic contour, the up and down pattern of a melody. In this reformulation of melodic perception can be seen a bit of psychoacoustics; the tone frequency, music analysis; the note in the octave and psychology and information processing; the pattern of a melody. Three areas which by this time had become integrated into the overall paradigm of cognitive research. This experiment was a direct result of the Attneave and Olson experiment described above. They

---

14. Dowling & Fujitani, 1970 op. cit.

15. Massaro, D.W., Kallman, H.J. & Kelly, J.L., The Role of Tone Height, Melodic Contour, and Tone Chroma in Melody Recognition, <u>Journal of Experimental Psychology: Human Learning and Memory</u>, 1980, Vol. 6, No. 1, 77-90

hypothesized that tone chroma would not be as functional in the recognition of unfamiliar melodies as it had been in the recognition of the NBC chimes part of the Attneave & Olson experiment.

Basically, the experiment took four six toned sequences which were original melodies to use as the original untransformed stimuli. They were called Transformation O. Then they transformed the melodies in four ways: Transformation PC preserved the original contour of the melodies without maintaining the exact pitch relationships; Transformation LT preserved the contour of the original melody and it preserved the size of the intervals but it cut down the range of frequencies by half and destroyed absolute pitch information; Transformation OPC, octave preserving contour, preserved the chroma and the contour information for each original melody but each note was displaced by one or more octaves; then finally, in Transformation OVC, octave violating contour, the notes of the melody were not only in different octaves but the direction of interval change in the original tune was disregarded. This violated the contour of the original melody but maintained the chroma of the original. The subject's task was to listen to the deformations of the melody and to respond by pressing one of four buttons whether the sequence sounded most like sequence 1,2,3, or 4. So in that sense it was a combination of an alternate choice and a matching response.

Massaro,Kallman & Kelly came to the following conclusions:

> "Performance on the OPC transformations was consistently much better than on the OVC and the PCS transformations. This result strongly suggests that tone chroma is functional in recognition of both highly familiar and recently learned melodies. Furthermore, all of the studies demonstrate that tone chroma information alone is not sufficient for accurate melody recognition. For chroma to be effective it must be accompanied by accurate contour information.
>
> We have argued that tone height, melodic contour and tone chroma may all contribute to the perception of a melody. How these characteristics are processed remains an important concern. For example, to say that tone height is important is not to say that the heights of the individual tones in a melody provide the cues necessary for melody recognition. Whether a given melody is sung by a bass or a soprano voice does not usually influence a listener's ability to recognize the melody. The important feature for melody recognition is that the relative heights of successive tones are maintained when a melody is transposed from bass to soprano."[16]

It becomes apparent in retrospect that the early testers in music aptitude couldn't have seen these relatively newly identified aspects of perception in the time and context in which they were working. The phenomena was there but unrecognized because no one was designing experiments which would reveal this kind of data. There were some who had the seeds of these ideas germinating. For instance, Drake's inclusion of a

---

[16]. Massaro, Kallman & Kelly, 1980, op. cit. p.89

lapse of time between tones in his retentivity test foreshadowed the use of delayed pitch comparisons in cognitive researchers years later.

To illustrate the similarities and differences of the old tests and the new experiments, I put together the chart entitled Comparison of Old Tests and New Experiments. (See next page) These "new" experiments which I have used for comparison purposes are in themselves old compared to research going on now, but my main goal here is to give a sense of the flow and change of perspective over the period from the old Atomist/Gestalt era into the Cognitive Science age. I think this comparative analysis does this.

I have designed my chart to illustrate where each of the tests and experiments fall in relation to the kinds of stimulus used and the kinds of responses called for. Obviously, it doesn't give the reader an in depth view of every detail of the tests and experiments. What it does show is that the way tests and experiments are designed are a reflection of the theoretical background and intentions of the designer.

In the first column are the names of the testers. The first four are the old testers. The second five, separated by a double line, are the new testers. The second, third and fourth columns list different kinds

COMPARISONS OF OLD TESTS & NEW EXPERIMENTS

Tester:	STIMULI: Pairs of Melodies	Pairs of Tones	Pairs of Chords	RESPONSE: Forced Choice	Alternate Choice	Matching Patterns
Seashore Measures of Musical Talents	Tonal Memory	Pitch		Pitch	Tonal Memory	
Kwalwasser-Dykema Music Tests	Tonal Memory	Pitch(1 tone)		Tonal Memory Pitch(1 tone)		
The Drake Music Tests	Retentivity Musical Memory	Retentivity Interval Discrimination		Retentivity Interval Discrimination	Musical Memory Retentivity	
Wing Standardized Tests of Musical Intelligence	1) Phrasing 2) Rhythm Harmony Memory		Pitch Chord Analysis	Pitch Harmony	Memory Chord Analysis	
Pflederer & Sechrest	Pairs of Musical Phrases			Pairs of Musical Phrases (Same/Dif., verbal disc.)		
Attneave & Olson	Pitch/Medium					Pitch/Medium
Deutsch		Pitch Comparisons		Pitch Comparisons		
Dowling & Fujitani	Preserved Contour Linear Transformation Octave Preserving Contour Octave Violating Contour				Preserved Contour Linear Transformation Octave Preserving Contour Octave Violating Contour	• • • •
Massaro, Kallman & Kelly	Tone Height Tone Chroma Melodic Contour				Tone Height Tone Chroma Melodic Contour	

of stimuli used in the tests; pairs of melodies, pairs of tones, or pairs of chords. The fifth, sixth and seventh columns list different kinds of responses; forced choice, alternate choice and matching patterns. As you look down the horizontal sections, you can see the names of the subtests in each testers battery which fall under each type of stimulus and response.

As you can see from looking at the chart, pairs of melodies has far outstripped pairs of tones as the stimulus of choice among the new testers. In the responses, forced choice was more predominant among the old testers, but not necessarily among the new testers. These changes reflect some basic changes in the way researchers are thinking about what music perception is and how to measure it, and what to measure.

Deutsch is the only experimenter that I used who still was using pairs of tones as stimuli in the new Cognitive age. Even she was using them with an interpolated melody which was supposed to serve as a distraction from the task of comparing two tones and she based her hypothesis for her experiment on Gestalt principles. Over time you can see that the use of pairs of melodies (or melodic phrases) as the basic unit to test perceptual phenomenon has eclipsed the old Atomistic strategy of individual tones compared.

Looking simply at the language which

experimenters use to describe properties of pitch perception, you can see the radical changes in approach that the Cognitive Revolution brought to music perception research. But underlying that, even, is the subtle shift from Atomistic two tone analysis to Gestalt wholes in the analysis of tone sequences or melodies in the actual objects that experimenters are focusing on. The difference being that, like the Aztecs and the Catholics, Gestalt wholes are the same phenomena transformed as the "Top-Down or Molar approach" these days.[17]

Looking at what has been learned by the new researchers and comparing that to what the early testers were trying to do (see chart entitled Old and New Conclusions) it becomes obvious that these new concepts come from a whole new way of perceiving psychological reality; through a "Cognitive" lens. The real differences that have taken place have taken place in the larger world of psychology and filtered down to make new connections with music perception which are then reflected again in the world of psychology.

---

17. Gardner, H., The Mind's New Science, Basic Books, N.Y. 1985, p.114

The terms molar and molecular were terms used by Gestalt psychologists:

Koffka, K. Principles of Gestalt Psychology, Harcourt, Brace, N.Y. 1935

COMPARISONS OF OLD/NEW CONCLUSIONS

Tester:	CONCLUSIONS:	UNDERLYING ASSUMPTIONS:	THEORETICAL BACKGROUND:
Seashore Measures of Musical Talents	It is possible to build a profile of music aptitude from meas. of individual sensory data..(ie..disc. of 2 tones)	Musical abilities are individual. Characteristics do not overlap. Musical abilities are inherited not acquired.	Sensorial/Elemental (Atomist)
Kwalwasser-Dykema Music Tests	"	"	Sensorial/Elemental (Atomist)
The Drake Music Tests	Memory for melodic material is an important and valid criteria for musical aptitude.	Musical factors cannot be measured separately from the context in which they occur.	Gestalt
Wing Standardized Tests of Musical Intelligence	"	"	Gestalt
Pflederer & Sechrest	Development of musical ability is related to the development of ability to conserve musical properties.	There are stages of musical development which correspond to stages of cognitive development.	Piagetian Developmental Gestalt
Attneave & Olson	Pitch is a medium in which the same pattern may have different locations.	perceptual objects owe their identity to certain relational invariants of patterns or configurations.	Psychophysics Gestalt Cognitive Psychology
Deutsch	Melodic sequences of small intervals are more easily processed than those of large intervals.	Perception of musical objects occurs on parallel channels of neurons.	Neurophysiological research. Information Processing, Cherry & Broadbent's British empiricism. Gestalt
Dowling & Fujitani	Precision of pitch and interval judgement encouraged by musical training is of little help in recognizing melodic contour.	The pattern of relationships among tones in a melody is what is important, and not their absolutely defined pitches.	Cognitive psychology, psychophysics, music theory. Gestalt
Massaro, Kallman & Kelly.	Two essential conditions for the recognition of melodies are tone chroma and melodic contour.	Tone height, melodic contour and tone chroma may all contribute to the perception of a melody.	Cognitive psychology, psychophysics, music theory. Gestalt

ALTERNATIVE METHODS OF ASSESSING MUSICAL APTITUDE AND PITCH PERCEPTION

In this chapter I will describe some alternative ways of looking at pitch perception theory and research. Particularly at some recent work of Jeanne Bamberger's and compare it to a more traditional style of research by Diana Deutsch. The chapter finishes with a description and analysis of a small follow-up study I did based on Deutsch's experiment.

Right from the beginning of experimental psychology there have been alternative perspectives about how to do research in the psychology of music and alternative theories as to how people hear music. In 1895 Billroth was proposing that musicality is more a characteristic of the central nervous system than a function of sense organs. He also suggested that there may be different kinds of musicality and that some of those kinds may be based on the type of music being played or listened to. (Billroth, 1895)[1]. Von Kries put forth the theory that musicality can be divided into two types, receptive and productive.(Von Kries, 1926).[2]. Revesz

---

[1]. Wer ist musikalisch?, Billroth, T., E. Hanslick, Pub. 1895

[2]. Musicality and Prognosis, Some factors related to success in school music situations, Holmstrom, L.G. Uppsala, 1963, Almqvist & Wiksells

characterized the Gestalt view of perception of pitch in the following way when speaking of relative pitch:

> "...regional pitch represents the special case of a generally valid organizing action of the mind, which is of decisive importance not only for the arrangement according to intensity, quantity, volume of sensory objects, but also for that of abstract ideas, such as numbers, let us say. This principle is one of the foundations of ordered thinking."[3]

Mursell carried on the Gestalt view of perception in his criticisms of the Seashore tests and in his book:

> "Musical relationships do not depend upon the physical properties of the sound wave or the physical action of the ear, but upon the integrating, organizing and selecting activity of the mind. And to regard the harmonic series as a determiner of basic musical effects is to fall into a primary psychological error, the error of attempting to explain an experienced whole in terms of the sum of its parts."[4]

Following Piagetian theories of conservation, Pflederer (1967) identified five types of music conservation:

"1. Identity- when thematic material maintains its essential characteristics across various permutations.

2. Inversions- in which an inverted simultaneous or successive interval is recognized.

3. Metrical groupings- in which meter recognition and discrimination are maintained despite changes in

---

Boktryckeri Ab. p. 20-24.

[3]. Introduction to the Psychology of Music, Revesz,G., U. of Oklahoma Press, 1954, p. 103-104

[4]. The Psychology of Music, Mursell, J.L. W.W. Norton & Co. NY 1937 p.55

note value distributions within measures.

4. Transposition- where a change of frequency level does not alter the perception of tonal configurations.

5. Augmentation and diminution- recognition that respective lengthening and shortening of a melodic passage's note values does not change the basic tonal relations." [5]

Pflederer continued pursuing research with these principles underlying the methodology developed. Pflederer and Sechrest published a paper in 1968 in which the data for the experiment was gathered in a similar manner to the clinical protocols used by Piaget:

"One hundred ninety-eight 5, 7, 9 and 13-year-old children were randomly assigned to experimental and control group training sessions. The experimental group listened to and discussed the first phrase of "America" in which the above[6] deformations had been made. For the control group, phrases from songs found in the basic elementary music series were used. These phrases represented slow and fast tempi, melodic movement in steps and skips, low and high pitch, multiple sounds, minor mode, and instrumental timbre.

After a training session each subject in that session met individually with an experimenter for thirty minutes for the tasks. Through a series of questions an attempt was made to determine whether or not conservation

---

[5]. Psychological Foundations of Musical Behavior, Radocy, R.E. & Boyle, J.D. Chas. C. Thomas, Springfield, Il. 1979 p.301

[6]. The deformations consisted of changes of instrument, tempo, harmony, mode, rhythm, contour, interval, and no change.

principles were evident."[7]

The responses of the children were divided into three groups within each age group; non-conservation, intermediate, and conservation. An example of the three responses can be found in the description of the harmony deformation part of the protocol:

> "The addition of harmony caused many subjects to hear more than one piano and even other instruments. In this deformation when perception centered on the harmony the tune was perceived as different...
>
> Seven-Year-Olds:
>
> Nonconservation: Different. They were using more fingers than the first one. (Were the songs the same in any way?) No.
>
> Intermediate: That's different because they played the piano. They played the piano with the other instruments. (Were they the same in any way?) Yes, the song was, but, um, some of the music in the middle wasn't because, um, the piano was playing.
>
> Conservation: If they took off the four, and then it would be the same.
>
> They had a low note after each note. They could be the same at the end if they took away the low notes.
>
> Well, it's just a little different. They had another new note after the other one. (Was there anything the same about them?) Not exactly. Just the notes if they didn't put the new....it would be the same if they didn't put the chord in it. Yeah, I think it

---

7. "Conservation-Type Responses of Children to Musical Stimuli", Pflederer, M. Sechrest, L. Council for Research in Music Education Bulletin, 13, 19-36, 1968

would be the same."[8]

The data was assembled into five different categories of verbal description: (A) Method/medium of performance; (B) Simple description; (C) Affective reaction; (D) Imagery and; (E) References to rhythmic movement. Category B, Simple description, in the harmony deformation elicited the most responses. The authors concluded that in terms of this experiment's usefulness to early childhood education and for the developing musical awareness of children, they should be given the vocabulary to be able to classify and label musical stimuli:

> "Responses denoting a kind of situational imagery indicate that children are able to relate personal experiences to music. As fascinating as these responses are, they also pose provocative and disturbing questions for the music educator. Why is it necessary for children who are studying music in elementary and junior high school classrooms to rely upon inexact terminology from other experiences to describe musical stimuli? "They had company" might be picturesque, but it is scarcely an adequate response for a 13 year-old to make to a chordal accompaniment.
>
> Why aren't children given the correct vocabulary as musical concepts are introduced? If a 7-year-old can identify a "contraction" and label it as such in language arts, why can't he also identify minor mode with it's proper label? " The second one got tired" conveys an original image, but is it really a substitute for the term "minor mode"? We do not believe that children's propensity for creative and picturesque speech should be discouraged, but we do believe that children should be given the vocabulary

---

8. Pflederer & Sechrest, op. cit. 1968 p.24

to facilitate classifying and labeling musical stimuli".[9]

At Harvard's Project Zero, Nelson Goodman's theory of symbols[10] provided a framework for some of the initial research done there into the arts and learning. In a report done for HEW in 1972,[11]. a description of Bamberger's experiments with children's conservation of rhythm and pitch is preceded by the following:

> "This question reflects a general view of perception as not simply a passive taking in (of) a stimulus, but an active organizing process in which the listener discovers or constructs his own coherence by spontaneously and/or deliberately coding the stimulus (not necessarily verbally) through various implicit means which can be described as categorical systems or languages. Thus, the perception of music varies as a listener's available categorical systems lead him to seize on different aspects of the stimulus and also according to the degree to which the listener can process multi-dimensional relationships."[12]

More recently Bamberger has just completed a new book in which she describes and analyzes in depth her experiments with adults and children in tasks related to constructing melodies and rhythms, drawing things on paper to represent the melodies and rhythms and how

---

[9]. Pflederer & Sechrest, 1968 op. cit. p.35-36.

[10]. Languages of Art, Goodman,N. Hackett Pub.Co. Indianapolis, IN. (1968)

[11]. "Basic Abilities Required for Understanding and Creation in the Arts", Goodman,N., Perkins,D., Gardner,H. Harvard Project Zero, Graduate School of Education, Harvard University, Cambridge, MA 1972

[12]. Goodman, Perkins & Gardner, op. cit. 1972 p.24

people make meaning of what they are doing.[13]

Bamberger's approach to understanding and describing musical development was preceded and influenced by Piaget but contains significant differences in focus and perspective:

> "I argue that while the sense of movement from one "stage" to another is certainly important, the sense of "progress" in this movement may be less so. That is, more elaborated organizing constraints (like Piaget's "mental schemas") which typically develop later are not unequivocally better: We must ask, "better for what?" This is a question that Piaget answers only implicitly. With his focus primarily on development that leads to capacities for symbolic abstraction which he associates with scientific inquiry, he unequivocally equates a later stage of development with, for example, the mental construction and internalization of fixed reference structures in relation to which particular properties of phenomena can be differentiated, measured and classified, these classifications remaining stable in the face of changing context, or as Piaget says, "in spite of the route traveled."
>
> In contrast to this focus and its prevailing view of development, I argue that, at least with respect to musical development, the active organizing constraints associated with an earlier stage of development need not be simply <u>discarded</u> to be replaced by or even absorbed into a later one. Rather, if we look on development as a <u>cumulative</u> rather than, let's say, a displacement process, this cumulative process can be a source for developing our capacities to make multiple hearings of the "same" musical material as we focus now on one kind of relationship, now on another."[14]

The basic research question that Bamberger asks in her book is: "What are the circumstances that generate fundamental ontological shifts associated with

---

13. Bamberger, J. <u>The Mind Behind the Musical Ear</u>, Harvard University Press, Cambridge, MA, 1991

14. Bamberger, J. 1991, op. cit. p.12 Chap.1

perceptual/conceptual restructuring--how do we ever come to see/hear in a new way?"[15] In order to come to some understanding about this question, Bamberger proposes a number of concepts in her book that come directly out of her years of experimentation and observation of her participants doing her protocols. From her point of view, teaching, learning and research are closely intertwined. It is partly her years of extensive observations of participants in her experiments that gives her work such credibility and coherence to me. Bamberger's experiments differ from more traditional methodology in important ways:

> "...my experimental designs include rather open-ended task situations which are often closely related to the musical activities that generated the puzzles in the first place. Further, the tasks are rich in possibilities for the active participation of subjects and sensitive interventions by the researchers. The carefully recorded on-going work of subjects becomes, then, the material for study and analysis. Interestingly, it is rarely a question of whether a subject can succeed in completing a task because almost all can, but rather the particular characteristics of a participant's work. For instance, I am concerned with the decisions made along the way, strategies used, shifts in focus including shifts in what the participant may be seeing as the problem to be solved, as well as the final product."[16]

Bamberger's methodology doesn't fit the traditional models of how to design an experiment and that is one of its strengths. The phenomena she is looking at

---

15. Bamberger, 1991, op. cit. p.10-11

16. Bamberger, J. 1991, op. cit. p.2

can't be easily understood with traditional methods of measuring and analysis. Bamberger has written about her work in an unconventional way as well. She begins her book with a prologue in which she "suggests that knowing how may be different from knowing about." She defines "knowing how" as "sequences of motions that we internalize in carrying out familiar activities- these action paths become our most intimate ways of knowing a piece." And "knowing about" as being "those with formal music training","Knowing -how is often differentiated from knowing-about to the extent that we can talk about what it is we know how to do. But what may be at least equally important are the matches and mis-matches between the theory of a domain as expressed in its privileged languages and the operative know-how of experts as well as non-experts." She goes on to suggest that there may be many different ways to "hear" the same piece of music and that a "hearing is itself a performance, an active process of making meaning."[17]

She has set up the rest of the book in three parts, in part one she focuses on different hearings of simple rhythms. She illustrates and analyzes these different ways to hear through her descriptions of working with children in a classroom setting as they devise a way

---

17. Bamberger, 1991, op. cit. prologue to Chap. 1

for themselves to "put their rhythm on paper so they can remember it tomorrow". What emerges from her careful analysis of the drawings is a "typology that includes two global dimensions-- on one hand, developmental stages, and on the other, differing aspects of musical structure which are reflected in the drawings at all levels of development."[18] It is from this kind of analysis and observation that her main theories have evolved.

In parts two and three of her book, she describes first working with children and adults in building tunes on a set of Montessori bells and devising notations for describing their constructed tunes. Then she looks in depth at her work with one student, Jeffrey, whom she worked with on tune building tasks over a period of six months. Bamberger describes a typical experimental situation in the following way:

> "The experimental situation always includes two parts: First I ask participants to build a tune. I will say, for instance, "Can you build Hot Cross Buns (or Twinkle, Twinkle Little Star) with the bells?" And second, I ask participants to "make instructions so someone else could play [the tune] on the bells as you have set them up." As might be expected, the construction strategy and the resulting finished product, together with the subsequent invention of "instructions" for playing the tune, provide important clues to builders' inner strategies for making sense of the tune, the situation, and the task."[19]

---

18. Bamberger, 1991, op.cit. p.4

19. Bamberger, 1991 op. cit. p. 162

In addition to her experimental methods described above, I think it would be helpful to the reader to understand what Montessori Bells are and some definitions of terms Bamberger devised to describe different kinds of activities related to the bells.

Montessori Bells:

> "Montessori and Macaroni designed the bells to be used in their classrooms as one of many sets of objects which are known as the "sensorial materials." The bells are a rather extraordinary technological invention in themselves. Each individual mushroomed-shaped metal bell is attached to a wooden stem, with bell and stem, in turn, standing on a small wooden base. A complete collection of Montessori bells includes one set with bases painted white, and another set that includes matching pitches but with bases painted brown. As part of the "sensorial materials" the two sets are typically used by children to listen for and pair brown and white bells that match in pitch. The bells, which are free to be moved about on a table, are played by striking them with a small mallet or "dinger" as we called it.
>
> Bells are carefully tuned so as to play different pitches (a complete set of white bells includes all the pitches of the C-Major scale; a complete set of brown bells includes all the pitches of the chromatic collection from middle-C to the C above). However, unlike other pitch-playing materials, all the mushroomed-shaped metal bells look the same. Differences in pitch, then, are distinguishable only by actually playing on the bells."[20]

Having given the reader a sense of what Montessori bells are, Bamberger goes on to define categories of activities related to working with the bells:

"Bell-Pitch: An object (bell) which has the property, pitch-P (e.g., G-bell or C-bell).

---

20. Bamberger, 1991, op. cit. p. 161

Bell-Path: The spatial <u>arrangement</u> of bells on the table, including all the forms this takes in the course of constructing the tune.

Action-Path: The sequence of <u>actions</u> made on the bell-path in playing the tune.

Tune-Path: The sequence of pitches in the tune as represented in standard <u>music notation</u> or in the participant's <u>invented notation</u>.

Table-Space:

a. Work space: the area of the table occupied by the cumulating bell-path; the space where the work of construction takes place.

b. Search space: the area of the table occupied by the bells in the mixed array which have not yet been included in the cumulating bell-path."[21]

Felt-Path: "The bells are like stepping stones along the uni-directional action-path; they mark the route through the tune much as the claps mark the route through a rhythmic figure. I call Jeff's strategy of construction a "felt path" strategy by analogy with the felt path of a clapped rhythm. Constructing the tune and then playing it is also reminiscent of the familiar experience of marking the chronological occurrence of chosen landmarks as one walks, through time, along a path from here to there--next-next-next. Traversing the path again, these temporally experienced landmarks become an on-going, enactive description of the path's unfolding--the way one remembers it in going along."[22]

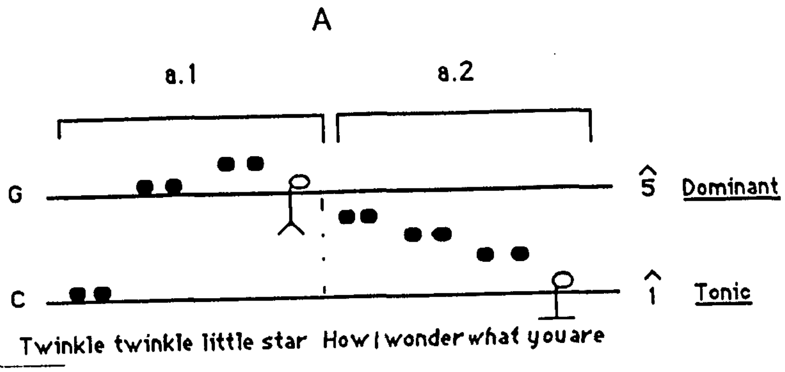
Bamberger prefaces her descriptions of actual experiments by giving a brief analysis of the musical structure of Twinkle, Twinkle Little Star. I am including it here so that my descriptions of her

---

21. Bamberger, 1991 op. cit. p.167

22. Bamberger, 1991, op. cit. p.251-252

subsequent work with Jeffrey will make more sense to the reader:



"The diagram shows the direction of pitch-motion (up and down) and also movement around two fundamental pitches--C and G. Blacks in the diagram indicate shorter durations, whites indicate longer durations. A white note with a indicates a boundary that is incomplete, a white with a indicates a boundary that is complete, closed-out.

Looking at the diagram, it is clear that the A section of Twink includes two distinctly different figures, a.1 ("Twinkle twinkle little star") and a.2 ("How I wonder what you are."). These two figures are not only different, but also strongly complementary to one another. The complementarity is expressed, first, by the differences in pitch-motion within each figure. Singing the first figure (a.1), notice that you go up in a leap, while in singing the second figure (a.2), you go down and you do so stepwise, the stepwise motion filling in the pitch gap left by the leap in a.1. Second and most important, complementarity is created by differences in "boundary conditions", specifically differences in structural functions at the boundaries. Singing a.1 again, you will hear and feel its boundary (on the word, "star") as incomplete, needing resolution; the second boundary (on the word "are") you will hear as complete thus resolving the incompleteness and the tension set up at the first boundary."

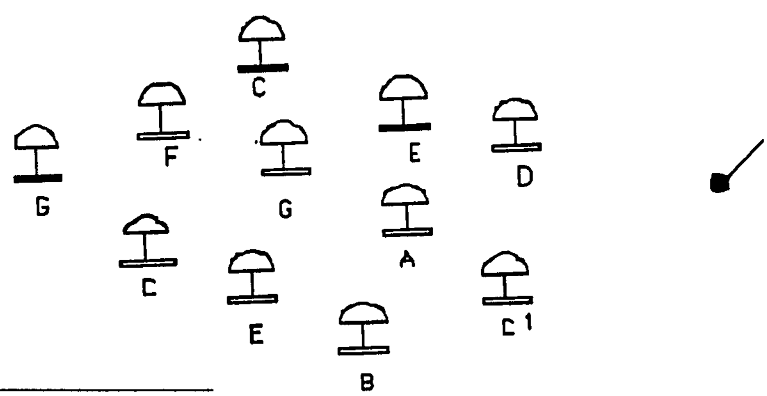
Having given the reader some overview of the musical structure of the tune, Bamberger goes on to describe

Jeffrey's constructions of Twinkle Twinkle Little Star. As the reader goes through the process of building tunes and drawing instructions to play the tunes with Jeffrey, the reader begins to experience the shift in focus from one kind of hearing to another that Bamberger is referring to. For example, Jeffrey goes through three stages in his constructions of Twinkle, Twinkle Little Star; first, a figural stage in which his construction of the tune and his drawing of it follow what Bamberger terms a figural strategy:

> "In short, using what I have called a figural strategy, Jeff invokes a mental setting in which each bell is recognized and given meaning as a particular and unique tune-event. And when he recognizes a bell as the object of his search and adds it to his cumulating bell-path, it marks a particular place as well as a unique structural function within the figure of which it is a member."[23]

Here is Bamberger's illustration of what Jeffrey did with the bells and the notations he drew to describe what he did:

He starts out with a "mixed array" of bells on the table:



[23]. Bamberger, 1991, op. cit. p. 243

"Notice that for this task I have included 11 bells in all-- the white bells include all the pitches of the C-Major scale and brown bells include matching C E and G bells."[24]

Jeff uses his usual figural strategy to build Twinkle.

Bamberger enumerates his actions in this way:

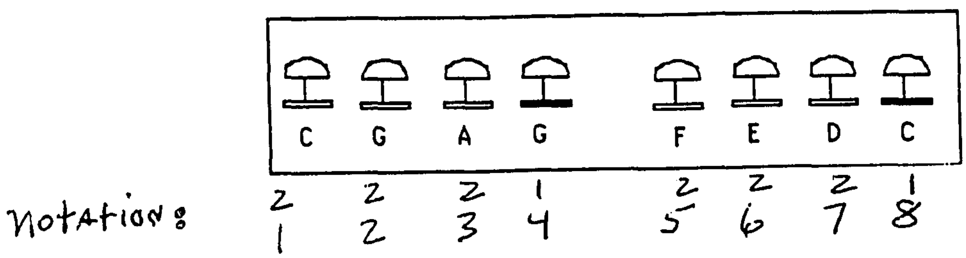
"1) Add a new bell to the cumulating bell-path in <u>order of occurrence</u> in the tune--next-next-next.

2) Add a found bell for next-in-tune to the <u>right</u> of the just previous bell in the cumulating bell-path.

3) Create a <u>context for search</u> by playing the tune-so-far, then test in the search space for a bell-pitch that matches the next tune-event.

4) There must be a <u>new bell</u> for each pitch-event in the tune even if a bell with the same pitch-property is already present in the bell-path."[25]

Jeff builds his prototypical figural bell-path:



Bamberger goes on to analyze Jeff's use of the bells to build Twinkle and points up the use of the brown bells for the second G and C (which match the white G and C bells) in the tune as typical of the figural

---

[24]. Bamberger, 1991 op. cit p.242

[25]. Bamberger, 1991, op. cit. p.243

strategy for building a tune:

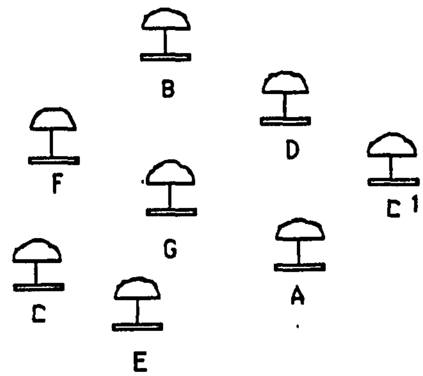
> "By introducing the brown bells, Jeff makes very clear again, that the meaning he gives to bells in space and events in time is determined by their particular situation along the route of the tune. Past and future exist only in the present moment; there is no comparing backwards or forwards in space/time. It is the particular sequence of tune-events and the position and function within that sequence that gives meaning both to the event and to the bell that instantiates that event. It is their situational properties rather than their fixed pitch properties that define them."[26]

Second, Jeff enters a transitional stage in which he faces a good deal of confusion in the process of trying to confront the disequilibrium brought about by trying to resolve the conflict between his figural representation of the tune and his attempts to integrate that knowledge into the task of building the tune with only one bell for each pitch in the tune. Up to this point, Jeff has had enough bells available in his array to use one bell for each event. As a result, he has been able to go from bell to bell following the tune events in order. The disequilibrium comes about when Bamberger removes the matching brown bells which gave Jeff the option of moving forward from one bell to the next through the tune. Instead, She has left only one bell for each pitch type in the array, all of them white bells. She then asks Jeff to play Twinkle with only the white

---

26. Bamberger, 1991 op. cit. p.249

bells.



At first Jeff has trouble building the tune without additional bells. Bamberger describes what Jeff does the next time he tries to build the tune a few days later:

> "A few days later I tried the same experiment again and this time Jeff accepted the challenge--perhaps because he was more comfortable with the task and thus more willing to risk trying something new. Arriving at the critical fourth tune-event, Jeff as usual started from the beginning of his cumulating bell-path (playing as far as "Twink-le twink-le lit-tle") to set the context for search. But then, instead of going into the search-space, he hesitated for a moment and turned back into his already built bell-path to search within it for a bell that would work for next-in-tune."[27]

Jeff continued to "switch back his action-path" over time and in this way was able eventually to play the whole tune on just the white bells. This was a significant moment in the transitional phase of Jeff's understanding of how to build the tune. Bamberger explains the significance this way:

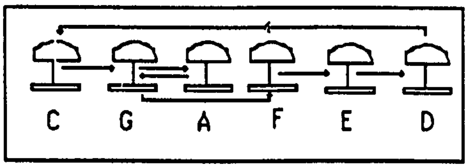
> "Most importantly, in making the switch-back in his action-path, bell-path, action-path and tune-path

---

27. Bamberger, 1991, op. cit. p.266

come apart: Jeff's bell-path (the sequence of bells in table-space), and the tune-path (the sequence of events unfolding in time) are no longer in spatial correspondence with one another; there is no longer a single, ordered series unified by a common chronology in space and time. It is Jeff's action path that serves as the means for coordinating the spatial incongruence of bell-path and tune-path."[28]

In the process of doing this task, Jeffrey switches his strategies so that instead of the placement of the bells on the table representing the figure, the order of actions on the bells comes to represent how to play the figure:



Transitional Bell-path and Action-path

"With his new transitional bell-path, Jeff's actions, his sequence of moves, act out the structure of the tune. So, the configuration of bells, an embodied description of the tune, has been replaced by an ordered sequence of actions, a procedural description of the tune...And instead of the spatial position of Jeff's bells serving as unique placemarkers for guiding and remembering his way, the chronology of his practiced sequence of actions must now do so."[29]

Still, Jeff had not come to a "formal" understanding of the musical structure in playing Twinkle little star. In order to see what would happen if she

---

28. Bamberger, 1991, op. cit. p.268

29. Bamberger, 1991. op.cit. p.277-278

intervened, Bamberger asked Jeff to "order the bells from low to high" Jeff ordered the bells into a tune. He thought she meant for him to "put them in order so they sound nice". Then she sang a C scale and asked Jeff if he could build that tune with the bells. He was able to quickly build the scale as a tune, next, next, next, just like he did with Twinkle. But he still did not understand the significance of the tune as a formal representation of a scale. Bamberger describes the results:

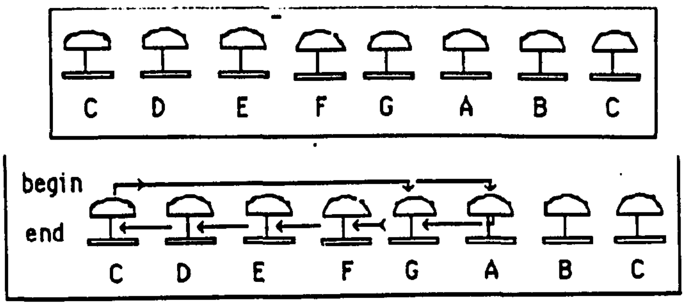
> "The result was, of course, a row of bell-pitches ordered consecutively from low to high, the whole series forming a major scale. But it is important to distinguish between, on one hand, a figural mental representation of this series, a "tune" and, on the other, what I will call a formal mental representation of this series. Formally, the set would be described as an ordered series in which any one bell-pitch in the series is both higher than the previous one (to its left) and lower than the subsequent one (to its right). With a formal representation, for example, a person would no longer need always to start from the "beginning" of the series--that is, the lowest pitch, rather the series could be built starting from any pitch."[30]

Bamberger had Jeff practice building the "scale-tune" for a few weeks and then asked him one day if he could play Twinkle on the "scale-tune". He had difficulty with this concept at first and continued to arrange the bells according to his transitional bell path. This marked the time when the two strategies, the scale-tune and Twinkle both existed for Jeff as two different

---

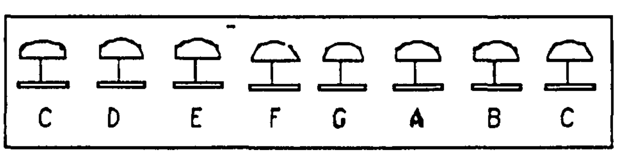
30. Bamberger, J. 1991, op. cit. p.306

kinds of tunes. It was only over time and practice building the scale-tune that Bamberger was able to help Jeff think about the scale-tune as not another tune but "as an instrument to play on--an instrument with a particular, stable geography."



playing Twink on his new instrument

It is in ordering the bells from low to high and recognizing how to use the "instrument with the stable geography" to play Twinkle twinkle little star that leads Jeff to understand that the names of the bells don't need to change in order for him to play Twinkle on it. The third stage of Jeff's construction and descriptions of the tune involve his discovery of formal music properties in the process of trying to describe on paper his new way of playing the tune on the bells:



The new instrument

[1] [5] [6] 5 [4] [3] [2] 1

and Jeff's instructions for playing Twink on it

> "In counting up on the bells in order to make his new notation, he tacitly labeled them, the bells, themselves, remaining as place holders within the fixed structure. In turn, Jeff, so-to-speak, "peeled" these invisible labels off of the objects, carrying the labels over one-by-one into paper-space. And in doing so, the number-names could gain an independent existence: They could be "picked-up", but still maintain their connection with the object from which they had been removed. Moreover, each time Jeff peeled off and carried over a number-name, the process left a trace; the cumulating row of numbers also gained a separate existence as a "notation" which referred to rather than imitating the sequence of objects that would play the tune... Indeed, Jeff's final notation can be called "formal" exactly because it depends on his invention of a rule system through which he can show the intersection between two structures each with its independently ordered sets of relations-- one, the bell series, which remains fixed in its order, the other, the sequence of tune events which is mobile depending on the tune involved, but with the second always described in terms of the first."[31]

Through the process of watching Jeff move from one kind of representation; the figural, to another; the transitional and finally to the formal, the reader's view of the tune changes as well and in this way it is possible for the reader to see and experience for themselves what Bamberger is trying to describe in her

---

[31]. Bamberger, 1991, op. cit. p. 326-327

theories.

One of the most unusual aspects of her book is the way in which she presents her material and then her theories which have come out of her work. First she describes the setting of the tasks, what motivated the tasks and the problem that was set for the subjects. Then she reproduces for the reader the exact narrative account of what actually happened in the setting; the subject's actions, what they said about what they did, and the drawings of the instructions for playing the tunes they constructed on the bells. Then she sets up an imaginary dialogue between two college-age students who are examining the tasks and results with her and coming to their own understanding of not only the subject's constructions and drawings, but also the whole process of doing the kind of research that Bamberger is teaching them to do. She has drawn the two imaginary student characters to epitomize the two predominant ways that participants in her experiments see/hear music; one is called "Met" for the metric approach and the other is called "Mot" as in motif or the figural approach. Bamberger characterizes the differences between Met and Mot in this way:

> "In the course of making descriptions, Met and Mot were able to discover that their differences often involved contentions over what each meant by "go together". For Mot, who prefers to "take-things-as-they-come", claps go together when, as a succession, they form a little bounded entity, a "figure." Met,

"jumping off" the internal time of his action-path, takes things "out-of-order" so as to select just those events that share some common property. For Mot, then, to "go together" means to form a class--like the class of all longs or all shorts."[32]

An important discussion takes place between Met, Mot and Bamberger after Jeff constructs the tune Hot Cross Buns with the Montessori bells. This all happens before he constructs Twinkle, twinkle little star, but it is a significant dialogue because it points up what is important about <u>context</u> in musical development:

"Met: ...why does he always play the whole tune, starting over again from the beginning every time that he goes in search of a new bell-- what you call "next-in-tune?" It takes so long, and he repeats everything so many times; especially later on as the tune grows. It seems pretty inefficient to me.

Mot: Now I've got a hunch about that.

Met: Yes, and...

Mot: Actually, it's about paths. Take a poem or a song that you've known all your life, they become like paths--you have to go with them from beginning to end. You can't just start in the middle; you have to start from the beginning and keep going. Wait, let's try it. Met, try to sing the Star Spangled Banner starting in the middle somewhere.

Met: O.K. (long pause; then Met sings, "What so proudly we hail..."). See, I did it. But I have to admit that I had to sing the beginning of the song to myself, first. Hmm, I never thought of that before. You mean Jeff has to play through the tune from the beginning each time to sort of catch up with where he left off?

Mot: Yes, I suppose you could put it that way.

Met: But that still doesn't explain why Jeff has to play every test bell 4 times when he's looking for the beginning of HA. Isn't <u>once</u> enough to hear if he's got it right or not?

---

32. Bamberger, 1991, op. cit. p.118

Mot: But you see, it's all part of the same thing.To tell if it is the right bell, Jeff first has to create the context, the environment, that leads up to it--like you had to sing the first part of the Star Spangled Banner to yourself before you could get the next part. He has to put himself on the tune-path, go as far as the bell-path goes, and test a possible continuation. And in order to really tell if it is the continuation, he has to act as if he's actually going on with the tune when he plays that test bell. And that means playing the bell 4 times as if it really were "one-a-penny". But if the test fails, like it did for the first white bell Jeff tries, then the context is broken and he has to start again--create the context all over again. So maybe it's not efficient as in fast, but it is efficient as in sure." [33]

The description of Jeff's needing to put the tune in context before being able to build the tune with the bells becomes significant when we look at the way more traditional tests of pitch perception and melodic perception are constructed. Which we will do when looking at an experiment by Diana Deutsch a little further on in this paper.

Finally, there are also Bambergers own comments about the dialogues with the students as well as information which she has included about specific musical aspects of the material the subjects are working with. It is through the reenactment of the protocols, the involvement of the reader in understanding the tasks, and the dialogues with the students that the reader comes to an understanding of Bamberger's theories.

---

33. Bamberger,J. 1991, op. cit. p.179-180

Bamberger's Theories

There are some central ideas in Bamberger's work that need to be defined in order to see her theories clearly. Because her theories have come out of her observations, she has developed a lexicon of terms to help describe particular features of her theories:

Simples:

"..refer to simple tunes and rhythms--those that we all sang or clapped as children. These actual tunes and rhythms that most of us learned in the natural course of growing up, I will call the <u>simples</u> of our culture."

Structural Simples:

"These are the small set of recurring pitch-time relations that, through cultural evolution, have come to be shared by all of our common folk & pop tunes, and by and large by all of the art music from at least Bach to Brahms."

Metric Simples:

"Consider for example, the regular marking off of time which is generated by the temporal relations among events in all of these common tunes-- the underlying pulse you tap your foot to... and consider the grouping together of these regularly recurring beats to form slower beats and their proportional divisions as these form faster beats."

Figural Simples:

"Figural simples are those shared pitch-time relations that serve to group together rhythmic and melodic events so as to form what we call phrases or figures...figural simples are characterized by their temporal symmetry or "balance".. a phrase that moves toward tension is most often temporally balanced by a phrase that moves towards stability."

A "Hearing":

"Thus, while a hearing may seem instantaneous, ineluctable, it is, in fact, a construction- an active play between the tacit, often unintended mental activities that we bring to bear and the yet to be organized stuff out there."

Multiple Hearings:

"What we hear depends on how we hear & how we hear is guided by where we focus our attention-- what features we choose or are able to attend to, how we segment and group them, and as a seeming inevitable result, what we see as the same and as different."

Figural & Formal Characteristics of Rhythm & Pitch:

"Expressions of similarity & difference among events result, then, from the specific kinds of possible features of the rhythm that each drawer chooses (or is able) to give precedence to: grouping of adjacent events into figures on one hand, and on the other, comparing measuring, and classifying of events according to their duration."

"Figural constructions and descriptions of melodies, like figural drawings of rhythms, are characteristically responsive to structural function and context-- in particular, the function of a pitch-event within the figure of which it is a member.

"In turn, those who make formal descriptions characteristically ignore function & context; two pitch events that share a common pitch property (two C's or two G's) will be labeled as the same in an invented formal "notation"...While we can speak of internal and relative pitch relations, such as "the melody goes up" or " gets higher"(just as rhythm gets faster or slower), in order to say how much higher, we must make use of a reference structure that has an absolute existence outside of the tune, itself. It is in terms of this outside fixed reference that we measure pitch distance and it is also with reference to this structure that pitch-names such as D or G gain meaning."[34]

Most people who study music may see the whole tune from the metric/formal view first, while most people who have not studied music may (or may not) see

---

[34]. Bamberger, 1991, op. cit.

the tune from the motific/figural view first. From Bamberger's point of view, regardless of whether the student is a Metric or Motific hearer, the goal of musical development is "to have access to multiple dimensions of musical structure- to be able to coordinate them & selectively choose among them, to change focus at will."[35]

In addition to the definitions of Bamberger's terms I have given already, I would like to give the reader a brief lexicon of musical terms which will be used in the analyses of Bamberger's and Deutsch's material:

Tonality: "Tonality is the organized relationship of tones in music. This relationship, as far as the common practice of composers in the eighteenth and nineteenth centuries is concerned, implies a central tone with all other tones supporting it or tending toward it, in one way or another...Tonality is synonymous with key. [It] is not merely a matter of using just the tones of a particular scale. It is more a process of setting forth the organized relationship of these tones to one among them which is to be the tonal center. Each scale degree has its part in the scheme of tonality, its tonal function.

Melody: A melody is any group of tones meant to be heard as a succession, the succession being organized in some way. Successive statement of the tones by a monophonic instrument ("melody instrument"), or by a singer, is of course a basic kind of melodic organization. More usually, however, when one thinks of melody it is a particular melody: one which has an identity, and an organization that is intended to be perceived.

Shape: The distribution of tones in a melody is marked by changes of direction, by range, by various high and low points, and by the variability of all of these as to where they occur in the phrase. Together

---

35. Bamberger, 1991 op. cit.

all of these aspects constitute the contour of the melody, an important determinant of its character.

Motive: The name motive is given to a short thematic unit,melodic or rhythmic or both, which is subject to repetition and transformation. A motive is thematic because it is recurrent and recognizable; at the same time, it is not usually an independent melody because it's characteristic appearance is as a constituent part of a melody. In all applications of a motive one expects to find variety as much as unity, and transformations as much as literal restatements. Yet the degree of resemblance, not the degree of variation, should be the criterion of whether or not a particular configuration of notes is motivic. If the transformation of the motive is so extensive that it is no longer recognizable as resembling the original, then it is no longer the motive but something else, perhaps a new motive; on the other hand, any resemblances among transformed motives will enable them to be considered as related.

Phrase: The phrase in music is comparable to the line in rhymed verse. The phrase shows a certain regularity in its number of measures, which is usually four or eight. It ends with a cadence, which is not a pause but something more like a breath that does not interrupt the flow of one phrase into the next. Most important, the phrase is perceived as a unit of musical thought, like a sentence or clause, and it generally implies that another phrase is to follow unless it shows a certain amount of finality. The phrase is what measures the beginning and ending of a melodic unit, as well as the point of departure for the next.

Phrase Structure: Phrase structure is one of the most important regulators of musical time. At the most immediate level, we perceive tones organized by rhythm and measured by meter, which is the perception of a regularly occurring pattern of strong and weak beats. At the most remote level, we perceive music organized into separate movements of different formal types, with a sectional structure characterizing each type, such as the exposition, development, recapitulation, and coda of the Classical sonata form. A hierarchy of phrase structure accounts for the various levels in between these time extremes. A small group of measures forms a phrase; a group of phrases forms a period or a subsection; half of a phrase may be a subphrase. An enormous variability in phrase types is to be encountered in music, which is why not all writers agree on what the different levels and categories of phrasing should be called. What is more important than an exact nomenclature is an appreciation of the different interpretations that may be made from different viewpoints.[36]

Getting back to her research question: "What are the circumstances that generate fundamental ontological shifts associated with perceptual/conceptual restructuring-- how do we ever come to see/hear in a new way?" In view of the descriptions of her work with Jeffrey, Bamberger's assertions at the beginning of her book that developmental stages could be cumulative as opposed to earlier stages being discarded to be replace by or absorbed into later stages, seems to be born out by the evidence she has presented to the reader in her book. What is particularly interesting to me is how her view of the way that people see music relates to other types of research being done in the field of pitch perception and music psychology. Given Bamberger's research, if tonality, shape and phrase structure, "a unit of perception of musical thought" as in the sense of beginnings and endings, where the root note or tonic is felt by the listener, where the accents fall, is the criteria by which people make sense of what they are hearing musically, measuring pitch perception in ways that don't take this structural function and context (as described in the

---

[36]. Piston, W. DeVoto M., Harmony, 4th ed. W.W. Norton, NY 1941-78, p.47-49

conversation between Met and Mot and in the definitions of musical terms above) into account would seem to be missing an important aspect. In my analysis of the following study by Diana Deutsch; "Delayed Pitch Comparisons and the Principle of Proximity," I will compare some of the issues and perspectives that Bambergers work has brought to bear on pitch perception with what Deutsch was trying to do in her study.

One of the ways in which I will analyze the differences between Deutsch's and Bamberger's work will be to look at what Deutsch is using as a stimulus in her experiment. In order to hear what Deutsch was trying to do with the stimulus in her experiment, I recreated the experiment with a few modifications of my own. Through my recreation of the tones generated from her description in her report of her experiment I was able to get an idea of what her subjects must have been listening to.

DELAYED PITCH COMPARISONS

Deutsch published a paper in 1978 entitled "Delayed pitch comparisons and the principle of proximity".[37] Deutsch describes the study as follows:

> "Subjects compared the pitches of two tones which were separated by a retention interval during which six extra tones were interpolated. The effects were

---

[37]. "Delayed pitch comparisons and the principle of proximity", Deutsch,D., Perception & Psychophysics, 1978 Vol. 23 (3), 227-230. (See Appendix II)

studied of varying the sizes of the melodic intervals formed by the successive tones of the interpolated sequence. It was found that error rates were lower when the interpolated sequences were composed of smaller melodic intervals; and it was argued that such sequences formed a more effective framework of pitch relationships to which the test tones could be anchored."[38]

Deutsch's reasons for doing the study was to further investigate the effect of manipulating the pitch relationships in the interpolated tones. Her idea was that the interpolated tones of smaller melodic intervals were more easily processed than larger intervals:

"It was reasoned that in listening to such a sequence we process not only the individual tones, but also the melodic intervals between them, which then provide a framework of pitch relationships to which the test tones can be anchored. So it was hypothesized that interpolated sequences forming melodic configurations which were more easily processed would be associated with enhanced levels of performance."[39]

The question that Deutsch was posing in doing this experiment was; are interpolated sequences composed of smaller melodic intervals associated with higher performance levels than interpolated sequences composed of larger intervals?

As plainly as I can put it, she seems to be measuring whether a subject in her experiment can more easily recognize a relationship between two tones when the six

[38]. Deutsch,D. 1978 op.cit.p.227

[39]. Deutsch, D. 1978, op. cit. p. 227

tones in between them are closer together (half tones) as opposed to farther apart (whole tones). I chose this experiment as the one to duplicate because I was interested in seeing whether a training period would have an effect on the scores. I decided not to duplicate her experiment in the sense of looking for the same results that she reported partly because my original objective had been simply to reproduce the stimulus that she described, not to duplicate the entire experiment. As a result, after duplicating Deutsch's procedures and stimulus I customized my experiment to look at the things that I was interested in. In Deutsch's experiment the procedure was as follows:

EXPERIMENT 1

"Procedure. In all conditions, subjects were presented with a test tone, which was followed by a sequence of six interpolated tones and then by a second test tone. They were instructed to ignore the interpolated tones, and to judge whether the test tones were the same or different in pitch. The subjects indicated their judgments by writing "S" (same) and "D" (different) on paper. The tones were 200 msec in duration and separated by 300-msec pauses, except that a 2-sec pause intervened before the second test tone.

Test tones. The test tones were all drawn from an equal-tempered scale (International Pitch, A= 435Hz) and ranged over an octave from Middle C to the B above. The frequencies employed (in hertz) were: C= 259, C#= 274, D= 290, D#= 308, E= 326, F= 345, F#=366, G= 388, G#= 411, A= 435, A#= 461, and B= 488. Within each condition, in half of the sequences the test tones were identical in pitch and in the other half they differed by a semitone. In half of the sequences where the test tones differed, the first test tone was higher than the second; and in the other half, the second test tone was higher than the first. All combinations of test-tone pitches were employed equally often in both conditions.

Interpolated tones. In both conditions, the interpolated tones were drawn from the same set as the test tones and were chosen at random from this set, except that no interpolated sequence contained repeated tones or tones that were identical in pitch to either of the test tones.

Conditions. There were two conditions in the experiment. In Condition 1 the interpolated tones were ordered at random, with the restriction that no more than three successive tones followed a unidirectional pitch change. In Condition 2, the interpolated tones were ordered monotonically; in half of the sequences, this order was monotonically ascending, and in the other half, it was monotonically descending.

There were 48 sequences in each condition, making 96 sequences in all. The sequences were arranged in random order, with no separation by condition. They were presented in groups of 12, with 10-sec pauses between sequences within a group and 2-min pauses between groups. Subjects listened to the entire set of sequences on two separate occasions, and the results were averaged.

Apparatus. Tones were produced as sine waves by a Wavetek oscillator controlled by a PDP-8 computer, and were recorded on tape. The tape was played to subjects on a high-quality tape recorder through loudspeakers.

Subjects. Fourteen undergraduates at the University of California at San Diego served as subjects for the experiment, and were paid for their services. The subjects were selected on the basis of obtaining a score of at least 8C% correct on a short tape designed as in Condition 1 (interpolated tones ordered at random)."[40]

Deutsch scored her tests by percentage of errors made by subjects. In her results and discussion she reports

---

[40]. Deutsch,D, 1978 op.cit.p.227-228

that the error rates were lower where the interpolated tones were ordered monotonically. She says this is in accordance with her hypothesis that interpolated sequences forming melodic intervals of smaller size would be associated with enhanced performance levels. Another possibility she considers is that because the interpolated sequences followed a unidirectional pitch change in the monotonically ordered sequences that this lends support to the hypothesis that unidirectional pitch sequences are more effectively processed than pitches which change direction.(Divenyi & Hirsch, 1975; Van Noorden, 1975) She adds that "this would also be expected from the Gestalt principle of good continuation."[41]

Actually, if you look at Deutsch's experiment from the perspective of Bamberger's experiments with Jeffrey, some interesting questions can be posited: For instance, if Bamberger is correct in her assertions that the fundamental ways in which we hear music as children happens first in terms of figural motifs rather than in formal/metric terms, the design of an experiment like Deutsch's may be failing to take important strategies that subjects use to understand what they are hearing musically (as pointed up by Bamberger's research) into account. Deutsch describes

---

[41]. Deutsch, D. 1978 op. cit. p.228

her reasons for doing her experiment the way she did as follows:

> "There is evidence from a variety of sources that tonal sequences are processed more effectively when they consist of melodic intervals of smaller size (e.g., Attneave & Olson, 1971; Bregman & Campbell, 1971; Dowling, 1973; Van Noorden, 1975); and indeed the early Gestaltists applied the principle of proximity to tonal sequences in the same way as they did to visual arrays (Koffka, 1935). So it was here reasoned that interpolated sequences composed of smaller melodic intervals would be associated with higher performance levels than interpolated sequences composed of larger melodic intervals."[42]

The way in which this design for an experiment is missing important aspects of pitch perception is that whether subjects are processing larger or smaller intervals while matching pitches is not related to the more fundamental question of how people come to perceive pitch in the first place. As Bamberger has pointed out:

> "It is the particular <u>sequence</u> of tune-events and the position and function within that sequence that gives meaning both to the event and to the bell [or tone in Deutsch's case] that instantiates that event. It is their <u>situational</u> properties rather than their fixed pitch properties that define them."[43]

In a sense, if you look at what Deutsch says about her hypothesis: "It was reasoned that in listening to such a (interpolated) sequence we process not only the individual tones, but also the melodic intervals between them, which then provide a framework of pitch

---

42. Deutsch, 1978, op. cit. p.227

43. Bamberger,J. 1991, op.cit.

relationships to which the test tones can be anchored." Right there she seems to have the beginnings of what Bamberger is already pointing to, (in the sense of the importance of relationship) but then when she goes on to say: "So it was hypothesized that interpolated sequences forming melodic configurations which were more easily processed would be associated with enhanced levels of performance." She's back into this traditional notion of "performance" in a test situation as being a passive observational "process" on the part of the subject who simply chooses between "same and different" and that this choice represents what is actually going on inside the subject. Bamberger's view of what is going on inside the subject is quite different. She says:

> "If hearing is indeed a process of instant perceptual problem solving, I need to ask, what are the sorts of processes that variously guide this perceptual problem solving? Putting the question this way, I obviously intend to suggest that what we casually call "the mind" is actively engaged in organizing incoming sensory material. And I want also to suggest that this is a generative process--that we are actively doing this organizing in real time as the sound/time phenomena is occurring "out there". But in saying that, I don't at all want to suggest that by "organizing" I mean some kind of "decoding" process, as if the incoming material has already been segmented, and these entities labeled or otherwise symbolically "encoded." Indeed, I will emphasize throughout the book that it is exactly because sound/time phenomena does not come already structured, but rather holds the potential for being structured that different hearings are possible."[44]

---

44. Bamberger, 1991 op. cit. p. 4-5

From this point of view, it is not so much the phenomena that Deutsch and other more traditional researchers are looking at that is different as much as it is how they go about investigating the phenomena. And as a result of that, or any one particular, style of investigation the description of the phenomena under investigation changes. Deutsch, Dowling and other more traditional researchers from the psychophysical end of the methodological scale of how to do research focus their efforts with instruments to measure based on the Weber-Fechner law of "just noticeable differences." As a result, the kinds of information which come out of these types of experiments are of a different order of magnitude than what Bamberger is looking at. The main difference which emerges between the two kinds of experimental designs is that on the Deutsch/traditional side, you have an experiment which tries to be "context-free", modeling itself after the sensory perception experiments of the earlier part of this century. This seems to have resulted in a focus on 2 stimuli comparisons in a lot of the experimental designs. Therefore, there is no tonal center or sense of key generated in the stimulus for this kind of experiment.

In contrast to that, Bamberger asserts that tonality; a melody's relationship and "loyalty" to a tonal center, it's "key"[45], is internalized at a very early time in a persons understanding of music. As a result, experiments which are done out of the context of tonality, as well as other elements of musical structure like accent, beginnings and endings of phrases, rhythm, contour and the like aren't going to give a researcher very much information about how a person makes sense of musical stimuli.

Another way of looking at the situation would be from the point of view of Bamberger's theories about multiple hearings. If context and tonal function play a predominant part in how people hear music, and if the ability to shift focus between one hearing and another is also a possibility, wouldn't that mean, then, that any particular analysis of test results from a study like Deutsch's would have to take into account alternative ways of hearing, which might alter the significance of the findings of an experiment like her's considerably? Deutsch ends her general discussion of the issues in her paper with more evidence to support her findings:

> "More general evidence is provided by the statistical distribution of melodic interval sizes in the music of various cultures. As pointed out by Dowling (1967), there is a striking tendency for the

---

[45]. Apel,W. Daniel, R.T., The Harvard Brief Dictionary of Music, 1961, Washington Square Press, NY

See also: music definitions, p. 109

frequency of occurrence of a melodic interval to be inversely correlated with its size. This has been demonstrated in the music of various primitive cultures (Merriam, 1964), Western cultivated music (Fucks, 1962; Ortmann, 1926), and recently in currently popular music (Jeffries, 1974). It seems plausible to suppose that such a distribution is based on an increasing inaccuracy or difficulty in the processing of melodic intervals as they increase in size. The present findings are in accordance with this line of reasoning."

Another possible explanation for the difficulty in the processing of melodic intervals as they increase in size might be that as they increase in size they lose coherence as melodic figures. In other words, the context would get lost. Wouldn't that explain just as well why subjects would have trouble "processing" them? Not to mention the possibility that "intervals" in and of themselves can change meaning and context, according to Bamberger. Even if they are not embedded in a melodic figure.

These are some of my thoughts on why the methodologies which are employed by more traditional researchers like Deutsch and others seem to be missing out on some essential characteristics of how people make sense of music. Maybe not in relation to the particular piece of research that this study is a part of but in terms of it's overall relevance to the larger questions; what is pitch perception? how do people make meaning of what they hear musically? or like Bamberger's research question: "What are the circumstances that generate fundamental ontological shifts associated with perceptual/conceptual restructuring-- how do we come to see/hear in a new way?." My basic question about the more traditional research methodologies is: are these kinds of research methodologies (like the one I described of Deutsch's) relevant to the phenomena being investigated? namely, how we hear and understand music? Based on what I've experienced in teaching and creating and reading in music, I would have to conclude that more effective research methodologies have yet to be devised, but Bamberger is definitely headed in the right direction. In order to examine more closely research methodologies like Deutsch's, what follows is an account of my recreation of Deutsch's experiment just described above. I recreated this experiment last fall using my music students at the school where I teach:

MY STUDY

Because the original impetus for my study was simply to hear what the stimuli generated for Deutsch's experiment sounded like, I was very careful to duplicate to the best of my resources and ability the pitches, durations, frequencies and wave forms of Deutsch's tones. I was fortunate in two ways; one is that I am very familiar with the Wavetek oscillator.

There are oscillators and oscilloscopes in almost every recording studio I've ever worked in. As a result, I had some idea as to what these tones were going to sound like. The second way that I was fortunate is that most midi systems are capable of producing the kinds of sine wave type tones that researchers only a short while ago had to invest many thousands of dollars into sophisticated and bulky equipment to produce.

Method

Apparatus:

The tones were generated by a Proteus/1 16 bit multi-timbral digital sound module midied to a Yamaha Kx88 controller using the solo synth factory preset edited to one octave only (Oct 1 (sine)) to minimize harmonics. I used a Macintosh SE computer with "Performer" software[46]. The tones were then transferred to high quality recording equipment and tape which was then played over JBL speakers to my subjects.

Procedure:

My procedure follows Deutsch's except for the length of the tones. I listened to the tones generated at 200 msec. and it was so short it was almost impossible for me or the professional engineer I was working with to

---

[46]. Performer, c.Mark of the Unicorn, Cambridge, MA

hear the pitch. As a result, I decided to lengthen the duration of my tones to 250 msec. At 250 msec. the tones were still very short but it was possible for us to discern the pitch of the tones. There was a 250 msec. pause between each of the interpolated tones.

Test tones:

My test tones were drawn from an equal tempered concert scale (A= 440 hz) and ranged over an octave from middle C to the B above. The frequencies employed in hertz were: C= 261.6, C#= 277, D= 293.6, D#= 311, E= 329.6, F= 349.2, F#= 369.8, G= 392, G#= 415.2, A= 440, A#= 466, B= 494. Except for the changes in frequency for the test tones, the other conditions listed under test tones in Deutsch's report were the same.

Interpolated tones:

The interpolated tones in the first two conditions or my study were as close to the same as Deutsch's as I could make them. The pauses between the test tones and the interpolated tones were a 1.75 sec pause after the first test tone and a 1.25 sec pause before the second test tone.

Conditions:

There were three conditions in my study. The first two were modeled as exactly as possible on Deutsch's two conditions. The third condition was designed to see if hearing melodies, both recognizable and unrecognizable, between the two test tones would have a significant effect on the subjects ability to remember the test tones. To that end, I used six interpolated tones, just like the first two conditions, but I altered the length of the tones and the interval distance to play the first six notes of six well known melodies and six notes of four original melodies, two major mode melodies and two minor mode melodies.

There were ten sequences in each condition making 30 sequences in all. Each sequence was separated by a 10 second pause and each group of ten were separated by a 2 minute pause. The sequences were not randomly placed. The first group were all random interpolated tones, the second group was made up of five monotonically ascending then five monotonically descending tones. In the third group of ten the well-known and original melodies were randomly placed. The well known melodies were: Jingle Bells, Camptown Races, Lullabye & Goodnight, How Gentle is the Rain, Twinkle, Twinkle, Little Star, and Mary had a little lamb.

Training:

My subjects listened to all three conditions of the test. Then over the period of a month and a half, ten of my subjects got training in how to remember a tone when your hearing something else: I had them hum the first tone to themselves while I played or sang a melody to them, and then had them guess whether the second tone I played was the same or different from the first. The rest of the subjects had no training. At the end of a month and a half, all of the subjects took the test again.

Subjects:

20 high school aged students were in my study. They ranged in age from 14 to 16. All of the students can be considered untrained subjects and they did the experiment as part of their regular music classes. There was no preselection of subjects except to the extent that they were all enrolled in some form of music class.

Analysis of the Data

All tests were scored based on the number of correct answers out of a total of ten in each of the three conditions and 30 for either the pre-test or post-test. Raw scores were tabulated for the pre-test, the post test, and for each of the conditions in each test situation. These scores were converted to percentages and comparisons of the total scores were analyzed based on performance of the trained group vs. the non-trained group. Also, the scores on the test were compared to grades in pitch perception which were given by the teacher's evaluation of performances over a period of one semester. The pitch grade was based on the subjects ability to sing or play an instrument in class. The range of grading was A=Excellent, B=Good, C=Average or D=Poor. An analysis of variance was done to understand the significance of the results.

Assumptions and Hypothesis

The conclusion that I came to after hearing the tones duplicated from the Deutsch experiment seemed fairly obvious on the surface,- this doesn't sound anything like music. Then the question arises, why doesn't it? Not so much from the perspective of Deutsch not designing her test to sound like music but from the point of view of why don't I process/hear these tones the same way I do music? This question came up again and again in doing this study. Especially when the subjects in the study would say to me afterwards, "that didn't sound anything like music". Once I had heard the stimulus (tones) that Deutsch was generating for her study, both in the sense of their timbre and their pitch, it became obvious to me that the timbre of the stimulus tone is not as important as what the tones are doing. By that I mean in the sense of musical context that I have described above in discussing Bamberger's theories. From that perspective, whether the tones are generated by a synthesizer, a piano, a sine wave or another instrument is not so important. What is important is the tonality, (as described in the definitions of musical terms on p.109) harmonic function and musical structure of what the participants in the experiment are listening to, how they make sense of that information and how I can get evidence of how they are making sense of it.

My attention then shifted to looking at how the interpolating tones worked in diverting the subject's attention from the test tones, as opposed to Deutsch's contention that the interpolated tones "provide a framework of pitch relationships to which the test tones can be anchored." My hypothesis was that if Revesz, Mursell, Piaget, Pflederer, Sechrest, Bamberger et al... were all saying that we first perceive music in terms of its functional properties, as in Bamberger's figural theory, then would my subjects have more trouble remembering the test tones if they were interpolated with material which more closely resembled music than the random tones and monotonically ascending and descending scales of the Deutsch study?. Even I and my engineer were fascinated by the boundary we crossed (perceptually) when we went from generating tones at 250 msec per tone to playing keys on a keyboard to generate the tones for the well known melodies. The minute the duration of the tones was a result of a hand on a keyboard, with all of the built in internalized musical framework that implies, we both experienced a perceptual shift of some sort from one kind of cognitive activity, which was more like a kind of "aural mapping"[47], to the much more familiar feeling for us, of listening to music. As a result, part of my assumptions and hypothesis is that what makes the tones bear any relation to music perception might be the phrasing and duration of the tones, and definitely what the functional properties of the tones are, where the boundaries lie, where the tonal center is, more than the greater or lesser degree of overtones generated. What does that say about what music is? What does that say about what studies like Deutsch's are actually measuring?

Results and Conclusions

There was a significant overall improvement of about 9.3% in scores between the pre-test and the post test. A comparison by condition of pre and post tests shows that the overall mean for each group of three was lower in condition 3 of each test than condition 1 of each test. To review; condition 1 was the random tones a semi-tone apart, condition 2 was the monotonically ascending and descending tones and condition 3 was the

---

[47]. "aural mapping" is a term I made up to delineate the difference between focusing attention on locating and comparing tones, as I experience in experiments like Deutsch's, and listening to melody or music, as I experience in experiments like Bamberger's or in composing and performing.

COMPARISON OF PRE- AND POST-TEST SCORES (n=20)

	A Cond.1	B Cond.2	C Cond.3	Pre-Test Total	% Variance B-A	% Variance C-B	% Variance A-C	S.D.
Pre-Test Scores	13.1	11.6	11.9	36.6	-11.5%	2.6%	10.1%	.123
Mean	.655	.58	.595	1.83				
				Post-Test Total				
Post-Test Scores	13.5	13.7	12.8	40	-1.5%	-6.6%	5.5%	.17
Mean	.675	.685	.64	2				
Variance Between Pre- & Post Test Scores.	3.1%	18.1%	7.6%	9.3%				

COMPARISON OF TRAINED PRE- AND POST-TEST SCORES (n=10)

	A Cond.1	B Cond.2	C Cond.3	Pre-Test Total	% Variance B-A	% Variance C-B	% Variance A-C	S.D.
Pre-Test Scores	6.3	5.8	6.0	18.1	-7.9%	3.4%	5.0%	.124
Mean	.655	.58	.595	1.81				
				Post-Test Total				
Post-Test Scores	6.1	6.4	6.2	18.7	4.9%	-3.1%	-1.6%	.196
Mean	.675	.685	.64	1.87				
Variance Between Pre- & Post Test Scores.	-3.2%	10.3%	3.3%	3.3%				

COMPARISON OF UNTRAINED PRE- AND POST-TEST SCORES (n=10)

	A Cond.1	B Cond.2	C Cond.3	Pre-Test Total	% Variance B-A	% Variance C-B	% Variance A-C	S.D.
Pre-Test Scores	6.8	5.8	5.9	18.5	-14.7%	1.7%	15.3%	.122
Mean	.68	.58	.59	1.85				
				Post-Test Total				
Post-Test Scores	7.4	7.3	6.6	21.3	-1.4%	-9.6%	12.1%	.127
Mean	.74	.73	.66	2.13				
Variance Between Pre- & Post Test Scores.	8.8%	25.9%	11.9%	15.1%				

melodies. Part of my hypothesis was that the melodic information in condition 3 would act as more of a distraction from the test tones than the other two conditions. The means of the pre-test do not bear this out in the overall population. The scores of condition 2 are 2.6% lower than condition 3 and 11% lower than condition 1. In the post test however, the condition 2 mean is 6.6% higher than condition 3. In the pre-test, subjects did best when they were listening to condition 1. Condition 3 came in second and condition 2 came in last. In the post-test, subjects did best on condition 2, then condition 1 and condition 3 was last. Between the pre and post tests the number of correct answers on the test rose by 18% for condition 2 but only 7.5% for condition 3. This would seem to point to the possibility that on a second hearing, the subjects were able to ignore the ascending and descending tones and more effectively match pitches in condition 2 than they were able to ignore the random tones of condition 1 or the familiar melodies in condition 3 to match pitches. Of the three categories, condition 2 scores improved the most in the post test. I don't know how much the factor of familiarity with the test has to do with the lower scores on the condition 3 post test, I would think that hearing the melodies the second time would make them easier to ignore for the untrained group and therefore the scores might be higher. For the trained group, I have to include the possibility that the training may have hurt their ability to attend to the pitch of the tones in all of the conditions. The other possibility for the generally lower scores of condition 3 is that subjects were in fact, distracted by the interpolated melodies and that melody is processed in a different way than the experimental tones of Deutsch's type of study.

My conclusions about the results of the overall scores are that: 1) There was an overall improvement in scores between the pre-test and the post test. This improvement is probably due to a combination of the music background the untrained group, (which I will elaborate on in the analysis of the trained and untrained group scores) familiarity with the test and the effect of training on the trained group. 2) The scores for condition 3 were lower than condition 1 for both the pre and post tests. For the pre-test though, the lowest score was in condition 2. Maybe the Gestalt theory of good continuation was stronger for ascending and descending tones than for the melodies in condition 3? This result though, is opposite Deutsch's results on the same condition. She reported her subjects as getting higher scores on condition 2 than on condition 1. From my understanding of her report, her hypothesis is that interpolated sequences which were more easily processed, (by which she means sequences composed of smaller melodic intervals,ie.. semi-tones, rather than larger intervals, ie..whole tones) would be associated with enhanced performance levels. The assumption being that hearing tones closer in pitch to the test tones would have an anchoring effect (her words) on the subject's ability to remember the test tones. Her results seem to bear her out. My hypothesis seems to take the opposite point of view that the function of the interpolated tones is to distract the subject from the test tones and that the more melodic properties there are in the interpolated tones, the more distracting they are. On a general level, my results seem to bear out my conclusions. and 3) There is a possibility that the fact that the scores for condition 3 (the melodies) were lower than condition 1 (random tones) was a result of the strong harmonic function and musical structure inherent in the melodies of condition 3. What I am saying is that the melodies acted as a stronger distraction to the test tones because of their strong melodic structure.

Analysis of Trained and Untrained Results:

There was an increase in correct answers for both groups. The scores for the untrained group from pre to post test showed an increase of 15%. The scores for

COMPARISON OF MUSIC GRADE TO
TRAINED PRE- AND POST-TEST SCORES (n=10)

	Pre-Test Total	Music Grade Total
Pre-Test Scores	18.1	27
Mean	1.81	

	Post-Test Total
Post-Test Scores	18.7
Mean	1.87
Variance Between Pre- & Post Test Scores.	3.3%

COMPARISON OF MUSIC GRADE TO
UNTRAINED PRE- AND POST-TEST SCORES (n=10)

	Pre-Test Total	Music Grade Total
Pre-Test Scores	18.5	35.5
Mean	1.85	

	Post-Test Total
Post-Test Scores	21.3
Mean	2.13
Variance Between Pre- & Post Test Scores.	15.1%

the trained group showed an increase of only 3.3%. Why such a big difference? One possibility that immediately came to mind was that somehow my training procedure hampered the trained subjects ability to accurately remember the test tones. Another factor is familiarity with the test and a third possibility is to look at the music grades for the two groups. The music grades are based on each student's performance on an instrument or voice in class over the period of a semester. The criteria for each grade is based on ability to stay on pitch, technique, rhythm and comprehension. When you compare the music grades to the scores on the post test for both the trained and untrained subjects, you can see that the average grade in music for the untrained group is 24% higher than that of the trained group. This is a factor which I did not account for when deciding who would be in the trained and untrained groups. For the most part, the untrained group is made up of singers. The trained group is made up of instrumentalists. This could answer a lot of questions about the poor response of the trained group. The singers are familiar with the idea of using their voices as instruments. The instrumentalists may not be. My training involved using your voice to maintain a pitch (humming to yourself). As a result, this may be why the training didn't help as much as I thought it would. The largest jump in scores were for the untrained group in condition 2 between pre and post tests. My hunch is that they were able to either use the ascending and descending tones as anchors for the test tones as Deutsch proposes or that they were more able to ignore the interpolated ascending and descending tones more effectively in this condition than in any other.

My conclusions about the analysis of the trained and untrained subject's performances are that; 1) the higher music grades of the subjects in the untrained group had a positive effect on their overall scores. 2) The fact that most of the untrained group was taken from the chorus and as a result had much more vocal training had a positive effect on their ability to remember pitch even without the training given to the trained group. 3) As a result, there is a possibility that the lower scores of the trained group were not a result of an ineffective training procedure but instead were a reflection of the relative lack of musical training in voice and possibly in general of the trained group as opposed to the higher level of training over a longer period of time for the members of the chorus who made up the untrained group. Which would imply that what Deutsch's study, which is what my study was based on, actually measures is how sophisticated a level of music training the subjects in her experiment display as opposed to any fundamental perceptual processes. Especially if what the choristers were doing was using the ascending and descending tones as "anchors" for matching the test tones as Deutsch proposed. That would imply to me that the subjects were able to hold an internal representation of some kind of formal musical scale which they compared the test tones to. And that would imply a high degree of musical knowledge on the part of the subjects. This would go along with my conclusions earlier in this chapter that the kinds of information that Deutsch is getting from her experiment is evidence of a higher level of musical understanding (ie..formal music theory students) than she may have intended when she designed her test. 4) There is also the possibility that because of the brief time involved for the training that changes which were supposed to be an effect of the training did not have much effect on the tests when compared to the level of a subjects general music and vocal training over a longer period of time.

Possibilities for the Future

Some questions and issues that I would like to pursue in the future about this experiment are: 1) Obviously, the next time I try to do an experiment of this type, I will check the background of my subjects very carefully, their grades in music, teacher's assessment of abilities etc.. before choosing the subjects for the research. One of the frustrating things about this kind of experiment is that I didn't find out anything about the way my students process pitch and to what degree they are able to recognize pitch that I hadn't already known from my own observations. A protocol like Bamberger's bell study would be much more rewarding in terms of getting a deeper understanding of the psychological processes at work. Another aspect of the experimental situation I'm interested in is longitudinal studies. Stanton and Gordon did exhaustive longitudinal studies based on Seashore and Gordon's tests.[48] But I would like to do longitudinal studies based on Piaget's, Bamberger's and Gardner's work. Based on Bamberger's work with Montessori bells, I would like to do an experiment using trained and untrained singers as well as trained and untrained musicians in a protocol set up like the one she described in her report on the Development of

---

48. Measurement of Musical Talent, The Eastman Experiment, Stanton, H. University of Iowa Press, Iowa City, Iowa 1935

Gordon, E. A Three-Year Longitudinal Predictive Validity Study of the Musical Aptitude Profile, Studies in the Psychology of Music, Vol.5, University of Iowa Press, Iowa City, Iowa, 1967

Musically Gifted Children. This would interest me because of her comments in this report about how trained instrumentalists use the bells.[49] Gardner's theories of multiple intelligences and symbolic development seem to me to offer the most comprehensive integration of many different kinds of research under the auspices of Project Zero. Following their lead, using observational techniques over time combined with more traditional testing methods at the beginning and end of a years worth of music study might lead to some interesting results;

2) There are many unresolved questions about how people process music as opposed to experimental tones. Deutsch described her interpolated sequences as forming melodic configurations in her report. From what I heard recreating the tones she reports using and from the reactions of my subjects after taking the tests, I don't know if I can call her random tones and monotonically ascending and descending tones melodic. The most visible proof of my skepticism being the lower scores that I got for condition 3 for both groups and overall. This gets back to the idea that a sequence of tones may or may not be meaningful depending on things like, harmonic structure, musical function, boundaries

---

[49]. Bamberger, J. The Development of Musically Gifted Children, Conceptions of Giftedness, Sternberg, R. (ed.) Cambridge Univ. Press

and multiple hearings.

3) A possible experiment to test this theory might include taking well known tunes like Happy Birthday and others and removing the pitch information while leaving just the durations in place and see if subjects can still guess what tune it is. And conversely, removing the duration and phrasing of notes leaving only the pitch to see if the tunes are recognizable. In Pflederer's theory of music conservation, augmentation-diminution is the type described as being when a subject realizes that the lengthening or shortening of notes doesn't change the relation of the pitches. This seems related to my questions about recognizing melodies. I would wonder how old a person would need to be to recognize a melody when played in 200 msec beeps with 300 msec pauses between each beep? (ala Deutsch's experiment) I found it difficult. Other similar experiments use tones as long as 500 msec. Other studies which point to new ways of looking at pitch and melody recognition such as Attneave and Olson's paper on pitch as a medium (1971), Massaro, Kallman & Kelly's theories regarding tone height, chroma and contour(1980), Cuddy's paper on hearing melodic patterns(1982), and others all deal with aspects of pitch taken out of their melodic context. But Bamberger's theories seem to address these issues more

directly and coherently than anything else I've read.

(4 The original reasons that music aptitude tests were developed were supposed to be to allow a teacher to assess the musical aptitude of a fairly large group of students in a relatively short time. This goal has never been reached by a music aptitude test as far as I can see. There are many other good reasons for doing research in music aptitude and pitch perception, which I have discussed in my qualifying paper and others have written about elsewhere, but the best method for a teacher to assess the musical abilities of her/his students still remains observation and the analysis and reflection that goes on in working with participants in a study like one of Bambergers over time. As a consequence, the Piagetian style protocols of Pflederer, Sechrest, Bamberger and the work of Gardner et. al. at Project Zero seem to me to point the way towards a deeper understanding of the underlying processes at work in pitch perception and other elements of music aptitude.

SUMMARY & CONCLUSIONS

Summary:

In this paper I have addressed the items from my Orals Memo in the following ways:

1. I have reviewed what I think are the pertinent theories which have influenced the design of the music aptitude tests that came out in the 1930's and 1940's and examined some features of those tests. This can be seen in the Literature Review and Description of Tests chapters of this paper.

2. I have looked at the follow-up studies that were done and what conclusions they came to about music aptitude tests and what criticisms they had for the tests. This can be found in the Follow-up Studies chapter of this paper.

3. I have looked at some features of the tests, particularly the way that they tried to measure pitch perception, and compared it to some new experiments to see what researchers have kept and what has changed in the design of experiments in music perception and aptitude over the years. This can be seen in the Description of Tests and Comparison of Pitch Subtest chapters of this paper.

4. I have looked at alternative methods of assessing musical aptitude and perception and compared some of the theories inherent in the new alternatives to some

more traditional methods of assessing pitch perception. I have also done my own follow-up study based on a traditional experiment and come to my own conclusions about what does and does not work in this kind of experiment. This can be found in the chapter on Alternative Methods of Assessing Musical Aptitude.

In addition to the foregoing, I have addressed my research question in the following ways:

Research Question:

> "How have notions about what music aptitude is and how to go about researching it changed, what theories have influenced that change, and what does that tell us about what we are doing now and where to go next?"

1. I have given a history of theories underlying the research done and described key elements of those theories and how they influenced the work of early testers and researchers in music aptitude.

2. I have examined four of the most influential tests developed to measure musical aptitude and compared the theories that influenced the development of these tests. I have also compared the methods and design of the four tests.

3. I have described some of the follow-up studies which came after the development the four tests mentioned above and compared and analyzed their results and criticisms of the tests.

4. I have compared elements of the pitch subtests from the early music aptitude tests with more recent

pitch perception and melodic perception experiments and analyzed the nature of the changes.

5. I have described alternative methods of assessing music perception and aptitude through the work of Pflederer and Sechrest, Bamberger and others. Then I compared Bamberger's work to Deutsch's and addressed what I see as the fundamental differences in perspective of the two.

In addition, I have included a description and analysis of a follow-up study I did last fall using the experiment of Deutsch's included in this paper as a model. I analyzed what did and did not work, I reported what I learned about designing tests from using Deutsch's experiment as a model for my own.

Conclusions:

Looking at my research questions again, what follows are my conclusions. I will present them to the reader in three sections:

1. "How have notions about what music aptitude is and how to go about researching it changed?"

First of all, the term aptitude is no longer in vogue amongst researchers in the field. The elements which together used to define "aptitude" in music; ability to learn, pitch rhythm, intensity, tonality, etc.. have been either abandoned or replaced by terms like; processing, perception, encoding, development, melodic contour, tone height, internal representation etc.. it seems to me that the whole concept of "aptitude" as a set of performance directives on the part of research and testing in education used to imply a sense of competition, our musical "grade" in the test of life. The new terms which refer to the same phenomena seem to point to a redirection of interest by researchers into the phenomena of pitch perception itself, without the underlying context of the competitive aspects of becoming adept at music. Unfortunately, focusing on the actual phenomena of pitch perception has not proved to have easy answers.

I have found that the research methodologies which were handed down from the founding fathers of experimental psychology; the laboratory setting, the preoccupation with measurement and trying to isolate factors and maintain "objectivity" are powerful procedural patterns which still have a lot of credibility and support amongst the general field of experimental psychologists. Of course, there are good reasons for these procedures to be used and they are very useful in many situations. There are traces of this pattern in almost every report of experiments in music perception, even when the researcher in question sees themselves as striking out into new methodological or theoretical territory. For instance, Attneave & Olson wrote in 1971:

"The basic fact that perception is relationally determined- that perceptual objects owe their identity to certain relational invariants of the psychophysical treatment of such commonly investigated continua as brightness, loudness, and pitch. A major defect of conventional psychophysics is its failure to take these consequences into account."[1]

Given this point of view, I would have thought that they would be interested in designing new ways of collecting the data, what kind of response was appropriate to finding "relational invariants", redesigning the whole concept of what constitutes a stimulus for such an experiment. Instead, even though from within the paradigm of what would be considered creditable research in psychophysics, they may have seen themselves as really going out on a limb. And even though the paper they did produce was very influential in helping to move research in pitch perception forward. In retrospect, given all of the possible options for what could be done based on the paragraph I just quoted from above, their experiment was actually very much designed to stay within the boundaries of accepted research methodologies in the field of psychophysics. The subjects heard two tones, the task was to match patterns, all of the tones were played on an oscillator to control any affects of

---

[1] Attneave & Olson, 1971 op. cit. p.147

overtones. There was a very elaborate system devised to sequence the stimuli:

> "The sequencing mechanism consisted of a phonograph turntable (rotating at 33 rpm) on which rested a cardboard disc extending beyond its edge. This disc was divided into 16 equal sectors, any combination of which could be cut away beyond the edge of the turntable to provide a photoelectric cam. Transparent sectors allowed a beam of light to strike a photocell; opaque sectors interrupted it. The photocell in turn controlled a circuit that switched back and forth between the two tones without temporal overlap, but without a noticeable interval between the alternating tones. Switching clicks were not altogether eliminated by the circuitry but were held to an unobtrusive level.."[2]

If perception is in fact relationally determined and owes its identity to certain "relational invariants", why not design an experiment that focuses on the relational properties of perception instead of focusing on keeping the stimulus free of any taint of human manipulation? In other words, if a subject is attending to some relational invariants in the foreground of doing a task, switching clicks as a faint background noise doesn't seem to be a significantly relevant factor in the phenomena under investigation. Nor does whether the sequences of tones are generated by a computer, or a person pushing some keys, or a turntable full of cardboard disks seem relevant. Attneave & Olson's real contribution to research, to my way of thinking, was in their placing pitch perception

---

2. Attneave & Olson, 1971, op. cit. p.149-150

in a new context: seeing pitch as a pattern which is transposable, " a medium in which the same pattern may have different locations."[3] My point then, is that the focus on "objective" means of measuring the phenomena in question is a leftover remnant from the paradigm for research that assumes understanding comes from devising tools to measure phenomena rather than from observation of the phenomena itself. You could argue that measurement is a form of observation but it is only one way of observing and understanding.

As a result, the question of what music aptitude is, is no longer a relevant question as posited. The question that remains is more like Bamberger's question: "What are the circumstances that generate fundamental ontological shifts associated with perceptual/conceptual restructuring--how do we come to see/hear in a new way?" This question seems to have more relevance to finding out how people hear, remember, learn and perform music than my original question.

2. "What theories have influenced that change?"

As I have noted before in this paper, the changes that have occurred in how music perception is researched have come from a combination of Gestalt theories, Cognitive Science theories and the research

---

[3]. Attneave & Olson, 1971, op. cit. p.148

methodologies of Piaget. For the more traditional researchers, like Deutsch and the psychophysicists and psychoacousticists, their research methodologies seem to have retained much of the core of procedures and assumptions handed down from earlier researchers like Helmholtz, Fechner and the like.

The ways in which I have chosen to illuminate the theoretical background of these experiments has been through an examination of the design of the experiments, the hypotheses of the researchers in doing the experiments, the results and what conclusions the researchers came to about their experiments and in what they said about their work, the language they used, where they were willing to go with their ideas.

3. "What does that tell us about what we are doing now and where to go next?"

In order to make clear the contrast between what Deutsch and other psychoacoustic researchers have to say about music perception and what Bamberger and the more alternative style of researchers have to say about music perception, I will present two examples:

1) In 1981 Deutsch & Feroe stated the following,:

> "It may generally be stated that we tend to encode and retain information in the form of hierarchies when given the opportunity to do so.....In considering how we form hierarchies, however, theories have generally been constrained by the nature of the stimulus material under consideration. For example, visually perceived objects are naturally formed out of parts and subparts. The hierarchical

structure of language must necessarily be constrained by the logical structure of events in the world. The attainment of a goal is generally arrived at by an optimal system of subgoals, and so on.

In this article we propose a model of how the observer represents the pitch sequences of tonal music in abstract form....However, our model differs from earlier ones in its basic architecture. In essence it may be characterized as a hierarchical network, at each level of which structural units are represented as an organized set of elements. Elements that are present at any given level are elaborated by further elements so as to form structural units at the next-lower level. It is further proposed that gestalt principles such as proximity and good continuation contribute to organization at each hierarchical level.

Before embarking on a formal description of the model, it should be noted that this concerns the representation of pitch information at the highest stage of abstraction, and that such information is assumed to be represented in parallel at lower stages also. At the lowest stage absolute pitch values are held to be represented, and interactions in storage that occur at this stage have been described elsewhere (Deutsch, 1975, in pressa). The next-higher stage is concerned with abstracted intervals and chords (Deutsch, 1969, 1978b). At the highest stage pitch information is further mapped onto a set of highly overlearned alphabets (Cuddy & Cohen, 1976; Cuddy, Cohen & Miller, 1979; Deutsch, 1977, 1980; Dowling, 1978; Frances, 1958; Krumhansl, 1979; Krumhansl & Shepard, 1979)."[4]

Deutsch sees this notion of hierarchy as being represented in stages of abstraction which are "paralleled at lower stages" and that "absolute pitch values" are held to be represented at the "lowest stage". It is interesting to compare this view to Bamberger's understanding of hierarchy and pitch

[4]. Deutsch, D & Feroe, J. The Internal Representation of Pitch Sequences in Tonal Music, Psychological Review, 1981, Vol. 88, No. 6, 503-522

perception:

> "If, for example, you are attending solely to "notes" as individual sound events and not at all to how they group together to form gestures or figures, then this would be not only a hearing at a particular level of attention, but a different kind of attention, as well....What emerges, then, are different aspects or dimensions of structure and it is important to make a distinction between them. For instance, hearings may differ with respect to level of attention within the structural hierarchy, on the one hand; and, on the other, a hearing may be figural/functional in its focus in contrast to a focus on formal properties. And, as with the children's drawings of rhythms, hearings may certainly focus on a mix of aspects and dimensions--in fact, they are probably most often just that--"hybrids". For example, a hearing may include more than one level at a time, and a person may also shift his/her attention from one aspect to another. Thus to make these distinctions is to once again expand the universe of "possible hearings" and in doing so to broaden the terrain in which to peruse the presumed goal of musical development, namely the capacity to make multiple hearings."[5]

From Bamberger's point of view "hearings may differ with respect to level of attention within the structural hierarchy, that the kind of attention that we give to a figural grouping is different than the kind of attention that we give to an individual note, attention can shift between different aspects as well as include different levels." I think it is significant that Bamberger points out that attention at the level of individual notes may be a different kind of attention than attention at the level of figural grouping. In that distinction lies one of the most important elements of the difference between the work

5. Bamberger, 1990, op. cit. p.158-159

of Deutsch and the psychoacoustic researchers and Bamberger and Gardner and other alternative researchers. Because Deutsch and others are focusing on "structural units made up of sets of elements paralleled on lower levels" It seems that they may be missing one of the really significant points of Bamberger's view of pitch perception; the shifting of attention between, and the ability to attend to, different aspects of structure and different kinds of hearing. Bamberger's view points up the fact that if there are different aspects of structure that are attended to and at the same time different kinds of hearing to be attended to, that would have definite consequences in how a researcher would design an experiment involving perception of music or pitch or rhythm. One way to go is in the direction of more observational techniques like Bambergers. I am not sure that delving further into "absolute pitch values" at the lowest stage is really going to lead to an understanding of phenomena which occur at the level of "figural groupings" amongst the youngest subjects. Actually, "absolute pitch values", in and of themselves, from a musicians point of view, are fairly meaningless.

Questions to be Explored:

The most intriguing question which I am left with after doing this paper is; how to formulate questions which will generate research methodologies which will bring about a clearer and deeper understanding of any particular subject. Particularly, an understanding upon which new theories can be developed? It seems that in spite of my consuming passion to make meaning of musical experience, my sense of curiosity has led me to wonder about how new methodologies for doing any kind of research come about? This is actually a very important question from my point of view because it seems pretty clear from this paper that the questions asked in research generate the answers to specific questions but may not necessarily contribute to understanding the phenomena in a new or better way. Not to mention developing theories about the nature of a given phenomena. The narrower the focus of the question, the more isolated and out of context the results seem to be.

On the other hand, if the questions are too broad in focus, how can you tell if your results are due to the situation you set up for your stimulus or due to unrelated factors? In his book Science and Subjectivity, Scheffler explains the relationship between ideas, imagination and theory:

> "Theory is not reducible to mere fact-gathering, and theoretical creation is beyond the reach of any mechanical routine. Science controls theory by credibility, logic, and simplicity; it does not

> provide rules for the creation of theoretical ideas. Scientific objectivity demands allegiance to fair controls over theory, but fair controls cannot substitute for ideas. "All our thinking" said Albert Einstein, "is of this nature of a free play with concepts; the justification for this play lying in the measure of survey over the experience of the senses which we are able to achieve with its aid."
>
> The ideal theorist, loyal to the demands of rational character and the institutions of scientific objectivity, is not therefore passionless and prim. Theoretical inventiveness requires not caution but boldness, verve, speculative daring. Imagination is no hindrance but the very life of theory, without which there _is_ no science."[6]

The distinction which is important to me in this explanation has to do with being able to distinguish between what the "institutions of scientific objectivity" really constitute as opposed to traditional research methodologies which may hamper finding new solutions to old puzzles. Is there a way to use imagination and "speculative daring" in designing research methodologies and new theories without losing sight of meaningful objectivity? I believe that is what Bamberger is up to.

With that question as a basis, the particular details of what kind of methodologies should be pursued and which should be left behind in the study of music

---

[6]. Scheffler, I. _Science and Subjectivity_, 2nd. Ed., 1982 Cambridge: Hackett Publishing Co. App.B, p.146

Albert Einstein, "Autobiographical Notes," tr. Paul Arthur Schilpp, in _Albert Einstein: Philosopher-Scientist_, Ed. P.A. Schilpp, New York: Tudor Publishing, 1949 perception would be left up to individual researchers. My own investigations into pitch perception and melodic representation are more likely to follow the lead of ethologists, ethnographic research, Gestalt theories and Piagetian methods for collecting data. As a musician, the work of Bamberger seems to be finding the most meaningful connections between how I hear and learn in music and the larger world of psychological theories about such things.

One more point I would like to make about doing research in music perception is that if you look at the research done from a historical point of view, one thing not taken into account by most researchers is that our cultural norms for what is musical have changed considerably since the beginning of this century. Not only has the music changed, the values about music have changed and the technology for producing music has changed so much that a new definition of what a "performance" is would not be out of order. People in the commercial/pop world of the music business have been referring to songs and scores as being basically a different kind of work of art from recordings since about the 1970's. This information is based on my own experience as a "recording artist" in the music business for the last 30 or so years. With the addition of digital sampling the connection between a living person being the focus of a recorded piece of music and the final product is severed. What keeps recording artists in business is the overall population of record buyers desire to hear records they can relate to. If the music business could create a perfect environment for itself, there would be no musicians, no singers, no recording artists. There would just be records; units of product that people would pay money for. This may or may not be the trend of the future, but it certainly puts the issue of what is musical and how people hear music in a different perspective.

Especially the criteria for pitch perception, whether a person is "on pitch" or "off pitch" certainly doesn't have much relevance when it comes to rap music. In a modern recording studio environment, technology controls pitch, duration, timbre, rhythm, all of the traditional categories which used to characterize the concerns of the early music testers. The ability to make a composite mix of many different performances has for the most part replaced the need for the performer to have perfected technique. This did not come about as a value judgement; whether performers are bad or good, it came about because the technology of the recording industry and the expedient need of the music business to save time (read money) in the studio has far outstripped the cultural norms and expectations regarding the kinds and quality of musical performance in our society.

As a result of what I have described above, the other question that I see as being an important one in doing research into music perception is how do we now define musicality?. Is there a definition of musicality which is not culturally bound? and finally, does research into music perception as opposed to perception in all areas, still have relevance to the world of music education? I don't know at this point whether it is still relevant to the world of commercial and pop music and the recording industry. If I had my way, of course, being a "contemporary acoustic singer-songwriter" we would all go back to "community sings" and singing and telling stories around a campfire to express our art. Even then, the issue of pitch perception would not be as important as an artists ability to give a "meaningful" performance in my ideal world.

Finally, that is why the importance to me of research in pitch perception and musical representation is probably going to be of more benefit as a contribution to the world of education and the psychology of perception and learning in the long run than it will be to the music business.

Appendix I

APPENDIX I

Chronology of articles & experiments:

Kwalwasser,J. Dykema,P.Kwalwasser-Dykema Music Tests. Carl Fischer Inc. 1930.

Drake, R.M. 'Four new tests of musical talent' Journal of Applied Psychology, 17, 136-47 1933

Drake, R.M. 'The validity and reliability of tests of musical talent' Journal of Applied Psychology, 17, 447-458 1933

Seashore, Carl E.,Two Types of Attitude Toward the Evaluation of Musical Talent,Psychology of Music, McGraw-Hill,1938

Swift, F.F.A Correlation of Kwalwasser-Dykema Test Scores Earned by Siblings. Unpublished thesis, Syracuse University, 1940.

Rubin, H.K. A constant error in the Seashore test of pitch discrimination. Unpublished master's thesis. University of Wisconsin, 1940 (Abstract)

Taylor,E.M., "A Study of the Prognosis of Musical Talent" Journal of Experimental Psychology, 10, I-28 1941

Bienstock, S.F.,'A Predictive Study of Musical Achievement' Journal of Genetic Psychology, 61,135-45 1942

Mursell, J.L. Psychological Testing, Longmans, Green & Co. N.Y. 1947

Lundin, R.W., 'The development and validation of a set of musical ability tests' Psychological Monographs, 63:305, I-20 1949

Piaget,J.,The Psychology of Intelligence, Routledge and Kegan Paul, London, 1950

Manor, H.C., "A Study in Prognosis" Journal of Educational Psychology, 41, 31-50, 1950

Cooley, J.C.,A study of the relation between certain mental and personality traits and ratings of musical

abilities, Doctor's Thesis, Michigan State College, East Lansing,MI 1952 (DA 13:240)

Lundin,R.W., An Objective Psychology of Music, The Ronald Press, New York, 1953

Skinner, B.F. The Science of Learning and the Art of Teaching, Harvard Educational Review, 24. 86-97 1954

Holmes, J.,'Increased reliabilities, new keys and norms for a modified Kwalwasser-Dykema test of musical aptitudes', Journal of Genetic Psychology, 85, 65-73, 1954

Revesz,G. Introduction to the Psychology of Music, Norman Oklahoma: University of Oklahoma Press. 1954

Bentley,R.R., A critical comparison of certain aspects of musical aptitude tests. PhD. thesis, University of Southern California. 1955

Christy,L.J., A study of the relationships between musicality, intelligence, and achievement. Phd. Thesis, Music, Indiana University, 1956

Kyme, G.H., "Are Musical Tastes Indicative of Musical Capacity?" Journal of Research in Music Education 4, 44-51, 1956

Meyer,L.B. Emotion and Meaning in Music, Chicago: University of Chicago Press, 1956

Langer, S.K. Philosophy in a New Key: a study in symbolism of reason, rite and art, Cambridge, MA: Harvard University Press, 1957.

Lundin, R.W. What next in the psychology of musical measurement? Psychological Records, 1958, 8, 1-6.

Wertheimer, M. Productive Thinking, New York: Harper Torchbooks 1959

Cain,M.L., A Comparison of the Wing Standardized Tests of Musical Intelligence with a Test of Musicality by Gaston and the Drake Musical Aptitude Tests. Master's thesis, University of Kansas (Lawrence, Kansas) 1960.

Ferrell,J.W., A validity investigation of the Drake Musical Aptitude Tests, PhD Thesis, Music, State University of Iowa. 1961

Gordon, E., 'A study to determine the effects of training and practice on Drake Musical Aptitude Tests scores',Journal of Research in Music Education,9 (I), 63-74,1961

Heller, J.J., The Effects of Formal Music Training on the Wing Musical Intelligence Scores. Doctor's thesis, State University of Iowa, (Iowa City, Iowa) 1962 (DA 23:2936)

Heim,K.E., Musical Aptitude of Senior High School Students in Residential Schools for the Blind as Measured by the Wing Standardized Tests of Musical Intelligence. Master's thesis, University of Kansas (Lawrence, Kansas), 1963.

Wertz, C.B.,Relation of Changes in Musical Preference to Scores on the Wing Standardized Tests of Musical Ability and Appreciation. Master's thesis, University of Kansas, (Lawrence, Kansas). 1963

Fosha,R.L. A Study of the Concurrent Validity of the Musical Aptitude Profile. Doctor's thesis, State University of Iowa, (Iowa City, Iowa), 1964 (DA 25:5319)

Tarrell, V.V., An Investigation of the Validity of the Gordon Musical Aptitude Profile. Doctor's thesis, State University of Iowa,(Iowa City, Iowa) 1964. (DA 25:5326)

Raim,R.L., A Comparison of the Musical Aptitude Profile and the Seashore Measures of Musical Talent, unpublished master's thesis, University of Iowa, 1965

Chomsky,N., Aspects of the Theory of Syntax, MIT Press, Cambridge, MA 1965

Piaget, J. & Inhelder B., The Child's Conception of Space, New York, W.W. Norton, 1967

Brown,M.E., An Investigation of the Optimum Length of Musical Aptitude Profile Subtests. Doctor's thesis, University of Iowa,(Iowa City, Iowa) 1967 (DA 28:2706A)

Davis, D.S. & Boone, D.R. Pitch discrimination and tonal memory abilities in adult voice patients, Journal of Speech & Hearing Research 1967, 10, 811-815 (Psychological Abstracts 42:5943)

Gordon, E., Implications for the use of the Musical Aptitude Profile with college and university freshman music students. Journal of Research in Music Education, 1967, 15, 32-40.

Gordon, E., A Three-year longitudinal predictive validity study of the musical aptitude profile. University of Iowa Monograph, Studies in the Psychology of Music, Volume 5. Iowa City, Iowa: University of Iowa Press, 1967.

Harrington,C.J., An Investigation of the Experimental Version Primary Level Musical Aptitude Profile for Use With Second and Third Grade Students. Doctor's thesis, University of Iowa, (Iowa City, Iowa) 1967 (DA 28:2713A)

Gordon, E., A study of the efficiency of general intelligence and musical aptitude tests in predicting achievement in music. Council on Research in Music Education Bulletin, 1968, 13, 40-45.

Pflederer,M. & Sechrest,L. Conservation Type Responses of Children to Musical Stimuli,Council for Research in Music Education Bulletin, 13, 19-36, 1968

Long, N.,Establishment of Standards for the Indiana-Oregon Music Discrimination Test Based on a Cross-Section of Elementary and Secondary Students With an Analysis of Elements of Environment, Intelligence and Musical Experience and Training in Relation to Music Discrimination, Revised 1969, An unpublished research project to the U.S. Office of Ecucation, Research Project No. 7-E-027. Indiana University, 1969, Pp.iv,79

Gordon, E., Intercorrelations Among Musical Aptitude Profile and Seashore Measures of Musical Talents Subtests, Journal of Research in Music Education, Vol. 17, 263-271. 1969

Piaget,J. Genetic Epistemology New York, W.W. Norton, 1970

Kuhn, T. Structure of Scientific Revolutions, 2nd. ed. University of Chicago Press, 1970

Rice, J.A., 'An Abbreviated Gordon Musical Aptitude Profile With EMR Children' American Journal of Mental Deficiencies, 75 (I):107-8 Jul. 1970

Zimmerman, M.P. & Sechrest, L. Brief focused instruction and musical concepts. Journal of Research in Music Education, 24,25-36

Zimmerman, M.P. Percept and Concept: Implications of Piaget. Music Education Journal 56, February, 1970, 49.

Attneave,F. & Olson,R.K. Pitch as a Medium: a new approach to psychophysical scaling,American Journal of Psychology, 84,147-66. 1971

Gallagher,'A Study of; the Relationships Between the Gordon Musical Aptitude Profile, the Colwell Music Achievement Tests, and the Indiana-Oregon Music Discrimination Test.' Doctor's Thesis, Indiana University,(Bloomington, Ind.) 1971 (DAI 32:2728A).

Dowling, W.J. & Fujitani, D.S. Contour, Interval and Pitch Recognition for Melodies, Journal of the Acoustical Society of America, 49, 523-531.1971

Gordon,E.,Studies in the Psychology of Music,(Ed.)University of Iowa Press, Vol. VI-VIII,1970-1972

Deutsch,D. Octave generalization and tune recognition, Perception and Psychophysics, 11,411-412. 1972

Bamberger,J.Basic Abilities Required for Understanding and Creation in the Arts, Goodman, Perkins,Gardner,(Ed.) Harvard Project Zero final report for U.S. Office of Education, Chap.4. 1972

Indiana-Oregon Music Discrimination Test, (revision of Oregon Music Test, Hevner,K. 1934) Revision by Long,N. Midwest Music Tests, 1975

Goodman,N. Languages of Art, Hackett, Indianapolis. 1976

Kuhn, T. A function for thought experiments, The Essential Tension (Ed.) Chicago: University of Chicago Press, 1977.

Dowling,W.J. & Hollombe, A.W., The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. Perception & Psychophysics, 21, 60-64.1977

Bamberger, J. Music and Cognitive Research: Where do our Questions Come From; Where do our Answers Go? Paper presented at the American Educational Research Association, San Francisco, CA 1979

Bamberger, J. & Brofsky, H. The Art of Listening New

York: Harper & Row, 1979 4th edition.

Bamberger, J. The Development of Musically Gifted Children in Sternberg, Robert, (Ed.), Conceptions of Giftedness, Cambridge University Press.

Bamberger, J. & Schoen, D.A. The Figural-Formal Transaction. Cambridge: Division for Study and Research in Education, MIT D.S.R.E. Working Papers, WP-1

Massaro, D.W., Kallman, H.J.& Kelly, J.L. The role of tone height, melodic contour and tone chroma in melody recognition. Journal of Experimental Psychology: Human Learning and Memory, 6, 77-80.1980

Gardner,H.,Davidson,L. & McKernon,P. The aquisition of song: a developmental approach. In Documentary report of the Ann Arbor Symposium, Music Educator's National Conference, Reston, Virginia. 1981

Serafine, M.L., Piagetian research in music, Council of Research in Music Education Bulletin, 1981.

Shuter-Dyson,R. & Gabriel, C.,The Psychology of Musical Ability, (2nd. Ed.) London:Methuen 1981

Gardner, H. Art Mind & Brain: A cognitive approach to creativity, New York, Basic Books. 1982

Bamberger, J. Revisiting Children's Drawings of Simple Rhythms: A Function for Reflection-in-Action, Strauss, S. (ed.), U-Shaped Behavioral Growth, Academic Press, 1982

Deutsch,D.,The Psychology of Music,(Ed.) New York, Academic Press, 1982

Howard, V. Artistry: The work of Artists, Indianapolis, Indiana, Hackett, 1982.

Dowling, W.J., Melodic Information Processing and its Development, The Psychology of Music, Deutsch,D. Ed. Academic Press, 1982 p.413

Serafine, M.L. Cognition in music. Cognition: International Journal of Cognitive Psychology, 1983, 14, 119-183.

Dowling, W.J. Scale and Contour: Two components of a theory of memory for melodies. Psychological Review, 85, (4), 341-354.

Gardner, H. Frames of Mind: The theory of multiple intelligences, New York, Basic Books. 1983

Gardner,H.,The Mind's New Science,A History of the Cognitive Revolution, Basic Books, New York 1985 Chap.2

Bamberger, J. Logo Music for the Apple Terrapin, Inc. Cambridge, MA 1986

Hargreaves,D.J. The Developmental Psychology of Music, Cambridge University Press, Cambridge, 1986

Hilgard,E.R., Psychology in America,a historical survey,,Harcourt, Brace, Janovich, New York 1987

Sloboda, J.A. Ed.,Generative Processes in Music, Clarendon Press, Oxford, 1988, vi

BIBLIOGRAPHY

Attneave,F. & Olson,R.K. Pitch as a Medium: a new approach to psychophysical scaling,American Journal of Psychology, 84,147-66. 1971

Bamberger, J. The Development of Musically Gifted Children in Sternberg, Robert, (Ed.), Conceptions of Giftedness, Cambridge University Press.

Bamberger, J. Music and Cognitive Research: Where do our Questions Come From; Where do our Answers Go? Paper presented at the American Educational Research Association, San Francisco, CA 1979

Bamberger, J. & Brofsky, H. The Art of Listening New York: Harper & Row, 1979 4th edition.

Bamberger, J. Logo Music for the Apple Terrapin, Inc. Cambridge, MA 1986

Bamberger, J. & Schoen, D.A. The Figural-Formal Transaction. Cambridge: Division for Study and Research in Education, MIT D.S.R.E. Wkorking Papers, WP-1

Bamberger, J. Revisiting Children's Drawings of Simple Rhythms: A Function for Reflection-in-Action, Strauss, S. (ed.), U-Shaped Behavioral Growth, Academic Press, 1982

Bamberger,J.Basic Abilities Required for Understanding and Creation in the Arts, Goodman, Perkins,Gardner,(Ed.) Harvard Project Zero final report for U.S. Office of Education, Chap.4. 1972

Bamberger,J.The Mind Behind the Musical Ear, Harvard University Press, Cambridge, MA 1991

Bentley,R.R., A critical comparison of certain aspects of musical aptitude tests. PhD. thesis, University of Southern California. 1955

Bienstock, S.F.,'A Predictive Study of Musical Achievement' Journal of Genetic Psychology, 61,135-45 1942

Cain,M.L., A Comparison of the Wing Standardized Tests of Musical Intelligence with a Test of Musicality by Gaston and the Drake Musical Aptitude Tests. Master's thesis, University of Kansas (Lawrence, Kansas) 1960.

Chomsky,N., Aspects of the Theory of Syntax, MIT Press, Cambridge, MA 1965

Christy,L.J., A study of the relationships between musicality, intelligence, and achievement. Phd. Thesis, Music, Indiana University, 1956

Cooley, J.C.,A study of the relation between certain mental and personality traits and ratings of musical abilities, Doctor's Thesis, Michigan State College, East Lansing,MI 1952 (DA 13:240)

Davis, D.S. & Boone, D.R. Pitch discrimination and tonal memory abilities in adult voice patients, Journal of Speech & Hearing Research 1967, 10, 811-815 (Psychological Abstracts 42:5943)

Deutsch,D. Music Recognition, Psychological Review, 1969, Vol.76, No. 3, 300-307

Deutsch,D. Octave generalization and tune recognition, Perception and Psychophysics, 11,411-412. 1972

Deutsch,D. Octave generalization of specific interference effects in memory for tonal pitch, Perception and Psychophysics, 1973, Vol. 13, No.2, 271-275

Deutsch,D. Error Patterns in Delayed Pitch Comparison as a Function of Relational Context, Journal of Experimental Psychology, 1974, Vol. 103, No.5, 1027-1034

Deutsch, D. Generality of Interference by Tonal Stimuli in Recognition Memory for Pitch, Quarterly Journal of Experimental Psychology, (1974), 26, 229-234

Deutsch, D. Musical Illusions, Scientific American, 1975, Vol.233, No.4, 92-104

Deutsch, D. Two-Channel Listening to Musical Scales, The Journal of the Acoustical Society of America, Vol. 57, No.5, p.1156-1160, 1975

Deutsch, D. Disinhibition in Pitch Memory, Perception & Psychophysics, 1975, Vol. 17, (3), 320-324

Deutsch,D.,Delayed Pitch Comparisons and the Principle of Proximity, Perception and Psychophysics, 1978, Vol. 23 (3), 227-230.

Deutsch, D. Interactive Effects in Memory for Harmonic Intervals, Perception & Psychophysics, 1978, Vol. 24, (1), 7-10

Deutsch, D. Binaural Integration of Melodic Patterns, Perception & Psychophysics, 1979, Vol.25, (5) 399-405

Deutsch, D. Ear Dominance and Sequential Interactions, The Journal of the Acoustical Society of America, 67, (1) 1980, 220-228

Deutsch, D. The Processing of Structured and Unstructured Tonal Sequences, Perception & Psychophysics, 1980, 28, (5), 381-389

Deutsch,D.,The Psychology of Music,(Ed.) New York, Academic Press, 1982

Dowling, W.J., Melodic Information Processing and its Development, The Psychology of Music, Deutsch,D. Ed. Academic Press, 1982 p.413

Dowling, W.J. Scale and Contour: Two components of a theory of memory for melodies. Psychological Review, 85, (4), 341-354.

Dowling, W.J. & Fujitani, D.S. Contour, Interval and Pitch Recognition for Melodies, Journal of the Ac oustical Society of America, 49, 523-531.1971

Dowling,W.J. & Hollombe, A.W., The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. Perception & Psychophysics, 21, 60-64.1977

Drake, R.M. 'Four new tests of musical talent' Journal of Applied Psychology, 17, 136-47 1933

Drake, R.M. 'The validity and reliability of tests of musical talent' Journal of Applied Psychology, 17, 447-458 1933

Ferrell,J.W., A validity investigation of the Drake Musical Aptitude Tests, PhD Thesis, Music, State University of Iowa. 1961

Gardner,H.,The Mind's New Science,A History of the Cognitive Revolution, Basic Books, New York 1985 Chap.2

Gardner,H.,Davidson,L. & McKernon,P. The aquisition of song: a developmental approach. In Documentary report of the Ann Arbor Symposium, Music Educator's National Conference, Reston, Virginia. 1981

Gardner, H. Art Mind & Brain: A cognitive approach to creativity, New York, Basic Books. 1982

Gardner, H. Frames of Mind: The theory of multiple intelligences, New York, Basic Books. 1983

Goodman,N. Languages of Art, Hackett, Indianapolis. 1976

Gordon, E., A Three-year longitudinal predictive validity study of the musical aptitude profile. University of Iowa Monograph, Studies in the Psychology of Music, Volume 5. Iowa City, Iowa: University of Iowa Press, 1967.

Hargreaves,D.J. The Developmental Psychology of Music, Cambridge University Press, Cambridge, 1986

Heim,K.E., Musical Aptitude of Senior High School Students in Residential Schools for the Blind as Measured by the Wing Standardized Tests of Musical Intelligence. Master's thesis, University of Kansas (Lawrence, Kansas), 1963.

Heller, J.J., The Effects of Formal Music Training on the Wing Musical Intelligence Scores. Doctor's thesis, State University of Iowa, (Iowa City, Iowa) 1962 (DA 23:2936)

Hilgard,E.R., Psychology in America,a historical survey,,Harcourt, Brace, Janovich, New York 1987

Holmes, J.,'Increased reliabilities, new keys and norms for a modified Kwalwasser-Dykema test of musical aptitudes', Journal of Genetic Psychology, 85, 65-73, 1954

Howard, V. Artistry: The work of Artists, Indianapolis, Indiana, Hackett, 1982.

Kuhn, T. Structure of Scientific Revolutions, 2nd. ed. University of Chicago Press, 1970

Kuhn, T. A function for thought experiments, The Essential Tension (Ed.) Chicago: University of Chicago Press, 1977.

Kwalwasser,J. Dykema,P.Kwalwasser-Dykema Music Tests. Carl Fischer Inc. 1930

Kyme, G.H., "Are Musical Tastes Indicative of Musical Capacity?" Journal of Research in Music Education 4, 44-51, 1956

Langer, S.K. Philosophy in a New Key: a study in symbolism of reason, rite and art, Cambridge, MA: Harvard University Press, 1957.

Lundin, R.W. What next in the psychology of musical measurement? Psychological Records, 1958, 8, 1-6.

Lundin, R.W., 'The development and validation of a set of musical ability tests' Psychological Monographs, 63:305, I-20 1949

Lundin,R.W., An Objective Psychology of Music, The Ronald Press, New York, 1953

Manor, H.C., "A Study in Prognosis" Journal of Educational Psychology, 41, 31-50, 1950

Meyer,L.B. Emotion and Meaning in Music, Chicago: University of Chicago Press, 1956

Mursell, J.L. Psychological Testing, Longmans, Green & Co. N.Y. 1947

Massaro, D.W., Kallman, H.J.& Kelly, J.L. The role of tone height, melodic contour and tone chroma in melody recognition. Journal of Experimental Psychology: Human Learning and Memory, 6, 77-80.

Pflederer,M. & Sechrest,L. Conservation Type Responses of Children to Musical Stimuli,Council for Research in Music Education Bulletin, 13, 19-36, 1968

Piaget,J.,The Psychology of Intelligence, Routledge and Kegan Paul, London, 1950

Piaget,J. Genetic Epistemology New York, W.W. Norton, 1970

Piaget, J. & Inhelder B., The Child's Conception of Space, New York, W.W. Norton, 1967

Piston, W, DeVoto, M.,Harmony, 4th Ed. W.W. Norton and Company, Inc. New York, 1978.

Revesz,G. Introduction to the Psychology of Music, Norman Oklahoma: University of Oklahoma Press. 1954

Rubin, H.K. A constant error in the Seashore test of pitch discrimination. Unpublished master's thesis. University of Wisconsin, 1940 (Abstract)

Scheffler, Israel, In Praise of the Cognitive Emotions, Science and Subjectivity, 2nd. ed. 1982, Cambridge: Hackett Pub. Co.

Seashore, Carl E.,Two Types of Attitude Toward the Evaluation of Musical Talent,Psychology of Music, McGraw-Hill,1938

Serafine, M.L. Cognition in music. Cognition: International Journal of Cognitive Psychology, 1983, 14, 119-183.

Serafine, M.L., Piagetian research in music, Council of Research in Music Education Bulletin, 1981.

Shuter-Dyson,R. & Gabriel, C.,The Psychology of Musical Ability, (2nd. Ed.) London:Methuen 1981

Skinner, B.F. The Science of Learning and the Art of Teaching, Harvard Educational Review, 24. 86-97 1954

Sloboda, J.A. Ed.,Generative Processes in Music, Clarendon Press, Oxford, 1988, vi

Swift, F.F.A Correlation of Kwalwasser-Dykema Test Scores Earned by Siblings. Unpublished thesis, Syracuse University, 1940.

Taylor,E.M., "A Study of the Prognosis of Musical Talent" Journal of Experimental Psychology, 10, I-28 1941

Wertheimer, M. Productive Thinking, New York: Harper Torchbooks 1959

Wertz, C.B.,Relation of Changes in Musical Preference to Scores on the Wing Standardized Tests of Musical Ability and Appreciation. Master's thesis, University of Kansas, (Lawrence, Kansas). 1963

Zimmerman, M.P. & Sechrest, L. Brief focused instruction and musical concepts. Journal of Research in Music Education, 24,25-36

Zimmerman, M.P. Percept and Concept: Implications of Piaget. Music Education Journal 56, February, 1970, 49.

VITA

Penelope Nichols Rothe

1967-Present	Singer/Songwriter, Vocal Arranger, Musician, Recording Artist	
1978-Present	Vocal Coach, private instruction	
1979-1981	Antioch University West Los Angeles, California	B.A. June,1981
1982-1983	Graduate School of Education, Harvard University, Cambridge, Massachusettes	Ed.M. June,1983
1984-1991	Doctoral Candidate Graduate School of Education, Harvard University, Cambridge, Massachusettes	
1985-1986	Teaching Fellow Graduate School of Education, Harvard University, Cambridge, Massachusettes	
1982-1986	Vocal Coach Office of the Arts Radcliff College, Harvard University, Cambridge, Massachusettes	
1985-1986	Teaching Assistant Music Department, Harvard College, Cambridge, Massachusettes	
1989-Present	Music Teacher Midland School Los Olivos, California	