R E P O R T R E S U M E S

ED 016 517 PS 000 330

DEVELOPMENT OF A TECHNIQUE FOR IDENTIFYING ELEMENTARY SCHOOL CHILDREN'S MUSICAL CONCEPTS. FINAL REPORT.

BY- ANDREWS, FRANCES M. DEIHL, NED C.
PENNSYLVANIA STATE UNIV., UNIVERSITY PARK

REPORT NUMBER BR-5-0233 PUB DATE SEP 67
CONTRACT OEC-6-10-002
EDRS PRICE MF-$0.75 HC-$7.80 193P.

DESCRIPTORS- *MUSIC EDUCATION, MUSIC TECHNIQUES, *ELEMENTARY SCHOOL STUDENTS, CONCEPT FORMATION, *MEASUREMENT INSTRUMENTS, STATISTICAL ANALYSIS, STIMULUS DEVICES, CURRICULUM DEVELOPMENT, *FUNDAMENTAL CONCEPTS, *MUSIC, COGNITIVE MEASUREMENT, BATTERY OF MUSICAL CONCEPT MEASURES, PENNSYLVANIA

CURRENT LITERATURE RELEVANT TO MUSIC EDUCATION OF CHILDREN STRESSES THE IMPORTANCE OF CURRICULUM DEVELOPMENT BASED ON A CONCEPTUAL APPROACH TO MUSIC LEARNING. THERE BEING LITTLE UNDERSTANDING OF CHILDREN'S CONCEPTS OF BASIC MUSICAL ELEMENTS ON WHICH TO FOUND SUCH CURRICULUM DEVELOPMENT, AN INSTRUMENT WAS DEVELOPED TO IDENTIFY THE IDEAS REGARDING PITCH, DURATION, AND LOUDNESS WHICH ARE POSSESSED BY CHILDREN. TWO GROUP MEASURES ELICITING WRITTEN RESPONSES TO (1) WRITTEN STIMULI AND (2) MULTIDIMENSIONAL MUSICAL STIMULI WERE DEVISED. IN ADDITION, 2 INDIVIDUAL MEASURES REQUIRING (1) THE MANIPULATION OF ELEMENTARY SOUND PRODUCING INSTRUMENTS AND (2) OVERT MOVEMENT IN RESPONSE TO MUSICAL STIMULI WERE FORMULATED. AFTER PILOT TESTING AND DEVELOPMENT, THIS INSTRUMENT WAS EMPLOYED IN A STUDY OF FOURTH-GRADE CHILDREN. THE GROUP MEASURES WERE ADMINISTERED TO 429 SUBJECTS AND THE INDIVIDUAL MEASURES TO 214 SUBJECTS. DERIVED SCORES, STATISTICAL EVALUATIONS, AND CORRELATIONS WITH IQ AND READING SCORES (REPORTED IN 61 TABLES) DEMONSTRATE THAT, ALTHOUGH FURTHER REFINEMENT IS NEEDED BEFORE IT WILL BE USEABLE FOR PRACTICAL PURPOSES, THE INSTRUMENT IS ADEQUATE FOR RESEARCH PURPOSES. THE DEVELOPMENT OF THIS MEASURE HAS SUGGESTED SEVERAL AREAS IN WHICH THE NEED FOR FURTHER RESEARCH IS INDICATED. (JS)

PA 24

FINAL REPORT
BR ~~Project No~~. 5-0233
Contract No. OE-6-10-002

DEVELOPMENT OF A TECHNIQUE FOR IDENTIFYING ELEMENTARY SCHOOL CHILDREN'S MUSICAL CONCEPTS

September 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

DEVELOPMENT OF A TECHNIQUE FOR IDENTIFYING ELEMENTARY SCHOOL CHILDREN'S MUSICAL CONCEPTS

Project No. 5-0233
Contract No. OE-6-10-002

Frances M. Andrews
Ned C. Deihl

September 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

The Pennsylvania State University

University Park, Pennsylvania

Acknowledgments

The research project reported herein represents the cooperative efforts of many individuals who have made constructive contributions of their time and knowledge. Sincere appreciation is due the following: The United States Office of Education for supporting the research project; Abraham W. VanderMeer, Dean, College of Education, The Pennsylvania State University, for unfailing encouragement and expression of interest; Harold E. Mitzel, Assistant Dean for Research, College of Education, for expert advice on research design; Robert L. Lathrop, Chairman of the Graduate Faculty, College of Education, for assistance as a test and measurements specialist; and the many school music supervisors, directors, and administrators who cooperated in the field testing phases of the research.

Three graduate assistants, Grace E. Laverty, A. Peter Costanza, and Cathy J. Cobes, made a major and outstanding contribution to the project; in addition, John Peightel, graduate assistant, was helpful in his brief association with the project.

The competent services of Christy Morgan are gratefully acknowledged, particularly in recognition of her general help as project secretary and as typist of the final report.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

OVERVIEW OF THE STUDY

Introduction

Major authorities in the field of music education, in speeches, articles, and textbooks published in the past several years, have devoted considerable attention to the structurally based, concept-centered music curriculum. This reflects a general trend in the search for a more efficient and effective method of communicating knowledge, and also indicates a specific interest in the reexamination of music as a subject to be taught and learned under the aegis of the schools. Current literature reveals a search for a core of musical content that can be accepted as common to the many areas of music education (e.g., instrumental, choral, general music, appreciation, and theory classes) and is also appropriate as a common basis for the musical learning of all segments of the school population.

Music has been variously described as both an art and a discipline with proponents of both camps making strong cases for their views. It is both, and derives considerable strength in its curricular status because of this duality. However, a complicating aspect of the duality is an uncertainty regarding the nature and status of the art and discipline. Extensive research into the characteristics of these two facets is needed to provide a sound framework for the music curriculum.

The present study has investigated an aspect of music as a discipline in its attempt to develop a technique for identifying children's concepts of certain musical elements.

Interest in a conceptual approach to musical learning is relatively recent. Although current literature relevant to the music education of children repeatedly cites the importance of basic musical learning to the development of musical concepts (Chapter II), the curriculum for music in the elementary school continues to be based primarily on subjective, experiential action and judgments rather than on a foundation of empirical evidence.

The current interest in the conceptual approach to music has revealed a gap between the opinions of leading music educators regarding curriculum content and the identification of a core of knowledge concerning the nature of children's concepts of music. A survey of the literature identified few research studies directly related to this topic. Yet it is paradoxical to advocate teaching musical concepts if there is little basic understanding or identification of the ideas children possess regarding music. Thus the need for a technique to identify children's concepts of music is apparent.

Psychologists have indicated that the young child does possess many concepts, even prior to formal school experience. McDonald (32) writes, "The child who is beginning school has already developed a system of concepts and characteristic ways of perceiving and organizing the stimuli from his

environment" (p.133). Russell (44) states that a child by the age of three or four knows hundreds of concepts (p. 229). Concepts are important to the learning process and learning behavior in that they are the foundation of problem-solving strategies. Each concept, as it develops, becomes the keystone for further concept development; each concept also interacts with other concepts to form complexes and hierarchies of concepts. Concepts are the basis of discrimination, value judgment, and generalization.

Accepting the importance of concepts in organizing learning behavior, it follows that the child's musical growth will rely heavily on his understanding of the organization and interaction of the structural elements of music. Little agreement is apparent among those who have written on the subject regarding which musical concepts should be taught, or at what particular time (Chapter II). The basic attributes of any musical tone, however, include the dimensions of pitch, duration, loudness, and timbre (pp. 5 to 7); the goal of the present study has been to develop a technique for identifying children's concepts of three of these dimensions or basic elements, pitch, duration, and loudness.

Through the interrelationships of these elements and their numerous and complex extensions of melody, harmony, dynamics, and tonal color, the core of a multifaceted, spiral, cyclic music curriculum should evolve. A conceptual approach necessarily involves these basic elements.

Purpose

In view of the foregoing, the purpose of this study was to develop a technique for identifying elementary school children's concepts of pitch, duration, and loudness. This involved developing and evaluating measures of musical concepts. Past tests have dealt with areas of music aptitude, music achievement, and music appreciation, but none has explored the area of musical concepts, essentially cognitive in nature. It was believed that the proposed technique might reveal significant information regarding children's concepts of musical elements, and possess considerable potential as a practical measure of childr. 's growth in musical concept development.

Definitions

In this study it is assumed that conceptual understanding of musical elements involves recognition of the basic elements of music in terms of their function within the musical frame of reference.

The basic elements of music are pitch, duration, loudness, and timbre. Their physical manifestations are frequency of vibration, duration of vibration, intensity or amplitude of vibration, and complexity of wave-form. The psychological dimensions of sound are closely related to the physical dimensions, although there is not a one-to-one relationship of frequency to pitch, duration of vibration to length of sound heard, amplitude or intensity to loudness, and the wave-form to timbre or tone-color.

The term dimension as used in the present study refers to the musical elements of pitch (higher/lower), duration (faster/slower, longer/shorter), and loudness (louder/softer). Because of the complexity of the problem of identifying timbre concepts, the dimension of timbre was not included.

The term multidimensional frame of reference refers to the complexity of the musical sounds within which the subjects were asked to perceive changes or differences. In the Listening and Overt Measures, subjects were asked to make a judgment of the change within a single excerpt, or a judgment on the difference between two excerpts, within the complex sound of the symphony orchestra. The music utilized was characterized by variations in pitch, duration, loudness, and timbre, constituting the multidimensional frame of reference.

The term predominant change refers to the most obvious or most prominent change in the music. Musical excerpts were selected with obvious or predominant changes restricted to one dimension or two dimensions.

The Battery of Musical Concept Measures included four measures, each constituting a different stimulus-response combination. The Verbal Measure required written response to written (verbal) stimuli. The Listening Measure required written response to musical stimuli. The Manipulative Measure required a manipulation of the stimulus material by the subject and a response by playing or indicating the correct answer. The Overt Measure required both an overt-physical movement response and an oral response to musical stimuli. The Verbal and Listening Measures, constructed for administration in a classroom situation, were group measures. The Manipulative and Overt Measures were individual measures.

Each of the four measurement modes measured the subject's concepts of the dimensions of pitch, duration, and loudness. The measures are referred to as balanced measures because they contained an equal number of items for the measurement of each dimension.

The Verbal and Listening Measures, designated as the Written Measures, were combined to produce a written score for each dimension. The derived scores are discussed as Written Pitch, Written Duration, and Written Loudness. The Manipulative and Overt Measures, designated as the Nonwritten Measures, were combined to produce a nonwritten score for each dimension. The derived scores are discussed as Nonwritten Pitch, Nonwritten Duration, and Nonwritten Loudness. Phrases such as "verbal pitch," "manipulative loudness," or "overt duration" refer to the score derived from a single measure on one dimension.

Music-making instruments, as used in the Manipulative Measure, were devices or contrived simple instruments on which musical sounds (i.e., sounds having perceivable pitch, duration, loudness, and timbre) could be produced easily by striking or plucking. In this study such devices included simple classroom rhythm instruments and bells, metal plates and metal tubes of various sizes, shapes, and densities.

Limitations

The study was limited to the development of a technique for identifying elementary school children's musical concepts. No attempt was made to determine how children form musical concepts or to measure growth in musical concepts. Other limitations were:

1. The study attempted to identify children's concepts of music in terms of the musical elements of pitch, duration, and loudness. Although timbre was identified as a musical element, concepts of timbre were not included in this study.

2. Subjects used in the Pilot and Main Studies were limited to children enrolled in the fourth-grade classes of twelve Pennsylvania and one Maryland public school systems. Classes were randomly selected from schools reporting adequate space for the administration of the measures. Subjects were within the normal range of intelligence and hearing.

3. The musical stimuli employed consisted of musical examples involving changes in the dimensions being measured. The symphony orchestra was the medium used for the musical examples in the Listening and Overt Measures. Other instrumental ensembles, solo instruments, and vocal music were excluded.

CHAPTER II

REVIEW OF RELATED LITERATURE

Background

This chapter includes a discussion of the elements of music, a general discussion of concepts, a discussion of musical concepts, and a review of related studies. The discussion of elements of music is necessary to develop a meaningful presentation of the study; similarly, the term "concepts" requires an overview, since the researcher investigating an area involving conceptual learning or the identification of concepts is confronted with a variety of schools of thought and viewpoints, as well as definitions. The review of related studies presents the results of an examination of the research relating to the present study.

Elements of Music

Increasing attention has been given in recent books and articles on music and music education to the development of musical concepts, but there is no general agreement on the identification of the musical elements to be emphasized for concept learning, or how concepts relate to such elements. Certain authors mention melody, harmony, and rhythm (Sheehy, 47; Leonhard and House, 30); other authors include these three, but add form or organization as a fourth element (Bergethon and Boardman, 3; Pflederer, 41; Hartshorn, 21; LaRue, 29; Monsour and Perry, 37). Ernst and Gary (15) discuss melody, rhythm, harmony, and timbre as musical elements (pp. 17-39); but they include, among desired musical outcomes, understanding of the "component parts of music and the interrelationships that exist between melody, rhythm, harmony, and form" (p. 6). Hoffer (25) states: "In music there are concepts of rhythm, melody, harmony, dissonance, modulation, syncopation, phrase, timbre, and many more" (p. 125). Woodruff (57) mentions concepts of form, rhythm, progression, harmony, counterpoint, and "many other precise musical ideas" (p. 226). Palisca (38) speaks of pitch and rhythm as well as timbre, dynamics, tempo, duration, form, and style (p. 9). Hermann (24) includes mood, tone, rhythm, melody, harmony, and form as essential musical learnings. According to Thompson (51):

> In sound, motion, and design we have the true essentials of the musical language. . . . There is no music without these three attributes. . . . The language of music is the language of melody, harmony, time values, rhythms, and musical form. All these are only variations on the theme of sound, motion, and design (p. 24).

The references cited to this point are characterized by lack of viable categorization in that certain writers, covering what is to them familiar ground, either omit one or more of the major musical elements and relevant musical extensions, or list as elements certain musical phenomena that

characterize musical performance or derive from a musical extension of an element (e.g., phrase, mood) but are not in themselves an inherent part of music in its elemental state. Rather, they are part of the musicianship of the composer or performer.

A different approach to the problem is taken by certain music educators, experimental psychologists, and physicists who define the properties of tone, or the structural elements of sound, as the basic elements of music. Among these authors the nomenclature applied to the various elements differs (for example, duration designated as time, loudness as intensity, timbre as tone quality or tone color), but there is agreement that a tone has properties of pitch, duration, loudness and timbre. Schoen (46) states that "all vibratory motion has the four properties of frequency, amplitude, form, and duration" (p. 5). Both Culver (13, p. 41) and Bartholomew (1, p. 6), writing on acoustics of music, name three respects in which musical sounds may differ: loudness, quality, and pitch. These are perceptual or sensory dimensions, dependent on physical manifestations of the sound wave (stimulus) as well as on the auditory mechanism of the receiver. According to Woodworth and Schlosberg (59):

> Loudness is primarily determined by stimulus intensity, even though the ear is much more sensitive to some frequencies than to others. Pitch is even more closely tied to stimulus frequency, with intensity having some minor effects. The auditory mechanism is an amazingly efficient analyzer of both frequency and intensity in the important middle ranges, even though less faithful at the extremes (p. 336).

Woodworth and Schlosberg also explain the correspondence between the physical and sensory dimensions:

> To physical intensity corresponds the sensory dimension of loudness; to vibration frequency corresponds pitch; to the composition of a complex wave corresponds the timbre of a musical instrument and the vowel quality of speech (p. 324).

Moles (36) presents three dimensions:

> A pure, isolated sound is a sonic entity defined by three dimensions, which may be physical: amplitude (pressure in baryes = dynes/cm^2), frequency (cycles per second), and length (seconds); or they may be perceptual: loudness (decibels), pitch (octaves), and duration (log $\underline{t}$) (p. 12).

Zuckerkandl (60) and Lundin (31) each mention five dimensions of tone. Zuckerkandl names pitch, loudness, duration, tone color, and volume; Lundin discusses pitch, loudness, timbre, volume, and density.

Considering these and other opinions, the consensus seems to be that a musical tone has the qualities of pitch, duration, loudness, and timbre. In the words of Fleming and Veinus (16): "The wave frequency, or the number of vibrations per second of the tone, leads to the sensation of pitch, its time length to duration, its energy to intensity /loudness/, and its overtone structure to timbre" (p. 5).

Bernstein (4) discusses relativity in the subjective measurement of tonal characteristics:

> The term pitch refers to the relative highness or lowness of a tone. . . . The duration of a musical tone may be defined purely in terms of the time elapsing between its commencement and its cessation. In actual practice the ear never measures the absolute time interval. It is, however, continually making comparisons between the relative time values of different tones, and between the time values of groups of tones.
>
> Intensity is the measure of the loudness of a tone. As in the case of the two previously mentioned characteristics, our measurement of intensity is purely relative. . . . The loudness of a tone for any given pitch depends on the amplitude of the vibrations; i.e., the extent to which the sounding body moves from its normal position of rest as it vibrates to and fro (pp. 1-2).

For the present study the four elements of pitch, duration, loudness, and timbre were accepted as the basic structural elements of music. The measures developed, however, dealt directly with pitch, duration, and loudness; timbre was excluded from the study.

Concepts

An examination of the literature revealed many definitions for the term "concept." Woodruff (58) states:

> A concept is a relatively complete and meaningful idea in the mind of a person. It is an understanding of something. It is his own subjective product of his way of making meaning of things he has seen or otherwise perceived in his experiences. At its most concrete level it is likely to be a mental image of some actual object or event that the person has seen. At its most abstract or complex level it is a synthesis of a number of conclusions he has drawn about his experiences with particular things (p. 2).

Vinacke (54) defines concepts as "cognitive organizing systems which serve to bring pertinent features of past experience to bear upon a present stimulus-object"(p. 294). Harriman (19) states that a concept is a "mental activity which brings two or more situations, experiences, or objects into a relationship; also, the sum-total of past experiences brought to bear upon a given situation" (p. 80). Hunt (26) quotes Webster's definition of a concept as a "mental image of a thing formed by a generalization from particulars; also, an idea of what a thing is to be in general" (p. 1).

In connection with the consideration of concepts, the distinction between "concept" and "percept" should be discussed. Russell (44) indicates that a "percept is the awareness of present data rather than a memory or image of things past" (p. 66). He classifies a concept as a "more or less stable percept" (p. 68).

Berelson and Steiner (2) define perception as:

> The more complex process by which people select, organize, and interpret sensory stimulation into a meaningful and coherent picture of the world. . . . sensation shades into perception as experience goes from the isolated and simple to the complex interpretations characteristic of normal, ongoing awareness of the world (p. 88).

The implication is that concepts result from percepts, and are more stable. Russell (44) states that "concepts develop out of related perceptual experiences" (p. 117); Smoke (50) refers to concept learning as the "process whereby an organism develops a symbolic response which is made to the members of a class of stimulus patterns but not to other stimuli" (p. 8).

Royer (43) lists five phases of developing a concept: sorting, classification, ordering, abstraction, and generalization. Harris (20) states that the three functions which comprise the process of concept formation are the perception, the abstraction, and the generalization (p. 5). Woodruff's (57) three phases of learning a concept are (1) perceptual, sense organs collect stimuli; (2) conceptual, the brain organizes the stimuli; (3) applicatory, the concept is used (p. 221).

Several studies indicate that a subject may possess a concept but may not be able to verbalize it. Heidbreder (23) found that "concepts were often used with consistent correctness though the subject was unable to formulate them verbally" (p. 673). Hebb (22) states:

> A concept is not unitary. Its content may vary from one time to another, except for a central core whose activity may dominate in arousing the system as a whole. To this dominant core, in man, a verbal tag can be attached; but the tag is not essential. The concept can function without it . . . (p. 133)

Russell (44) indicates that in some experiments children show a clear understanding of a concept but are unable to verbalize it (p. 118). However, Chan and Travers (10), in a study in which they showed ambiguous visual displays simultaneously with labels either descriptive of the display or irrelevant, found that "meaningfulness and relevancy or appropriateness of the label attached to the stimulus may be a factor facilitating perceptual learning" (p. 65).

Concept formation involves comparing, discriminating, judging, organizing, abstracting, and generalizing. Formation of concepts involves the application of previously acquired knowledge and understanding to new situations, and the forming of useful and consistent inferences and conclusions.

Differentiation between percepts and concepts is difficult, for percepts form the background of concepts. Hunt (26) states that "neither the process nor situational differences between perception and conception are 'all or none' distinctions" (p. 6). The lines between percepts and concepts are fine ones, and the learner crosses back and forth from one to the other in the process of concept formation.

If a hierarchy is involved in the formation of musical concepts, one can speculate that it is somewhat less than discrete. It may develop in such a manner as this:

1. Any sound causes the listener to respond in terms of an aural sensation (e.g., dropping a book or closing a door is a noise stimulus causing the individual hearing it to respond in terms of an aural sensation).

2. Awareness of a different, highly organized sound, such as a musical sound, is a more complex and prolonged sensation. The listener must now be aware of a kind or classification of sound that possesses distinguishing characteristics from such sounds as those of the falling book or the closing door. This awareness might be said to be limited initially to the perception of a musical tone, a fragmentary stimulus. When perception of a musical tone, or a timbre, broadens to include that of many timbres (e.g., different voices and instruments), or when the perception of musical movement broadens to include that of rhythmic movements, it is at least close to constituting a concept.

3. Awareness and recognition of a musical entity, such as a song, a march, a dance tune, performed in a musical medium (e.g., solo violin, orchestra, piano) as an organized unit, involves many single discrimination percepts which interact to form higher order concepts.

4. Awareness of the specific musical factors of duration, loudness, pitch, and timbre are conceptual functions when recognized in terms of their particular operational behavioral function in a musical framework. For purposes of this study, such a framework is referred to as multidimensional.

Related Studies

A review of related research revealed few studies directly pertinent to the identification of children's basic musical concepts.

Deihl (14) investigated musical concept development of college students through two subtests: basic structure concepts (verbal stimuli) and performance quality concepts (aural stimuli). Utilizing regression analysis, he found a low, nonsignificant correlation between musical concept development and amount of performance experience; amount was determined by weighted questionnaire. The subtest of structural concepts correlated nonsignificantly with the amount of experience while the subtest of performance concepts correlated low but significantly with amount of experience.

Il'ina (27) studied musically backward preschool children in an attempt to determine "which elements of a melody find reflection in the auditory conceptions of the child . . . and to what degree motor vocal reactions participate in the formation of these auditory images." A conclusion was that motor reactions are important in differentiating auditory imprecisions (p. 710).

Pflederer (40), in a pilot study, investigated certain responses of five- and eight-year-olds to certain musical tasks embodying the Piagetian

principle of conservation. Each task represented an hypothesis that developmental trends in the conservation of the specific musical concept (meter, tone, rhythm) could be discerned through the administration of the task to children of different ages. In seven of the eight tasks the eight eight-year-old children were better able to conserve the concept in question than were the eight five-year-old children. A primary implication of this pilot study was that overt interaction of the child with the musical problem seemed to be essential.

In a later paper Pflederer cited findings of Brehmer and Rupp, as reported by Franklin (17). Brehmer found tonal thinking of children to be dominated by the function of the musical configuration as a whole; young children cannot think in terms of abstractions that deal with the various parts of a total melodic shape. Rupp found that children before age eight are not yet ready to listen both horizontally and vertically at the same time. His subjects seemed to perceive only the melody in harmonized music.

Simon (48), replicating Pflederer's tasks in a further study with six- to nine-year-olds, found a general improvement in performance with increasing age. Comparing visual conceptual tasks involving seriation and area conservation with musical analogues, he found that the musical tasks were more difficult than their visual analogues.

Simon found that pitch discrimination appeared extremely difficult, requiring not only the perceptual ability to discriminate between tones but also the conceptual understanding of the terms "high" and "low" in music. Auditory seriation was an almost impossible task for the large majority of subjects, even subjects who demonstrated a conceptual understanding of seriation in the visual domain. Only in the metric conservation task did any process of logical inference seem to be involved. Simon concluded that assessing the individual roles of perception and conception in the tasks is difficult.

Boekelheide (5) designed measures to assess the listening skills of eight- and nine-year-olds in areas of sensitivity to rhythmic and melodic movement, aural recognition of like and unlike phrases, and aural recognition of changes in the pitch level of phrases. She found high correlations between musical development, as measured by performance on her tests, and both general ability and academic achievement, as measured by standard intelligence and achievement tests. The mean scores obtained for the rhythmic response measure were higher than for the melodic contour and pitch discrimination tests in this battery. Boekelheide believed her tests identified the general range of low and high achievers in the listening skills but did not identify individual musical characteristics of subjects.

Williams (56) pointed out, among other difficulties that characterize studies in pitch discrimination with young children, the problem that naming pitches as "high" and "low" is an arbitrary convention which must be learned. He used piano for demonstrating this relationship to four- and five-year-old children and found all children able to learn quickly and demonstrate at the piano relationships of "high" as "upstairs" and "low" as "downstairs." At the conclusion of a short testing period during which children were asked to identify pitch changes in five patterns, however, some children were unable to replicate the "upstairs-downstairs" piano demonstration. From a subsequent study of the ability of four-year-old children to sing songs, he concluded that, even after a year of daily training in singing, some children lacked pitch consciousness.

Williams (56) found evidence that "the concept of relative loudness has become stabilized in practically all normal children by the time they are four years of age, in many cases even younger" (p. 17). In a study which involved having children reproduce a periodic beat by tapping, Williams found failures decreasing from 75 per cent at age three to 4.1 per cent by age six, with no failures at ages seven and eight. Williams' studies seemed to indicate that ability to discriminate and reproduce loudness and rhythmic changes develops earlier than pitch consciousness.

Sievers (56) used children from Grades 1 through 6 in a related study using Williams' tapping device. He asked children to reproduce, at three different speeds, two patterns: (1) a steady beat and (2) a longer note alternating with a shorter one of half the duration. He found that accuracy in reproducing the patterns varied with age of subjects and speed of beat. Younger children had difficulty reproducing either pattern accurately at slower speeds, with many children in the lower grades unable to reproduce the second pattern at any speed. He found an increasing per cent of correct responses through the first four grades, with all children above fourth grade correctly reproducing the patterns at any of the speeds used.

In another study, using seven one-measure rhythmic patterns, Sievers found an irregular development of the ability to reproduce the patterns; this development correlated positively with age. He also found that correlations between chronological age and rhythmic ability were considerably higher than correlations between rhythmic ability and intelligence. He concluded that rhythmic performance seems to improve somewhat with age but not with intelligence.

Riley and McKee (33), in a investigation of pitch and loudness transposition, found that of ninety-seven first-grade children only one failed to learn the loudness discrimination but twenty-four failed to learn the pitch discrimination with a 500-cycle difference. In a study with second- and third-grade subjects and adults (42), they found that second- and third-graders had difficulty in learning a 100-cycle pitch discrimination but easily learned the 500-cycle difference. All were able to learn the loudness task easily. Riley and McKee proposed that the more readily learned amplitude discrimination may be based on previously established mediating responses which the subjects bring to the experimental situation. They felt six-year-olds to be more familiar with the concepts "louder-softer" than with "higher-lower."

Jeffrey (28) investigated differences in response mode for tonal frequency discrimination learning with five-and-one-half-year-old children. He found that the words "high" and "low" lacked meaning for those children when applied to musical tones. His results seemed to indicate that differentiation of widely-separated pitches was learned more successfully when the subjects responded vocally or by playing the note on the piano than when they merely pressed a button on the left or right, producing no sound. All subjects showed inability to transfer from the training interval of three octaves and a fifth to the smaller interval of a fifth.

Bridges (7) studied the harmonic discrimination ability of children in Kindergarten through Grade 3. She found that with increasing age of subjects there was a gradual development in the ability to discriminate between appropriate and inappropriate harmonic accompaniments to songs. Another

finding was that the majority of children at any age are better able to discriminate harmonically with an unknown than with a known melody. These results seem to indicate that children's musical perception is influenced by past experiences with music.

Similar results were found in Ward's (55) study in the area of auditory perception. He concluded that individual differences in aural perception can result from a common stimulus, that aural perception originates within the hearer, and that threshold differences cause differences in perceptions of tone quality.

In his studies of children's auditory perception of musical sounds, Petzold (39) found that children improve on all dimensions with age. His longitudinal studies revealed that most children reached a plateau in these rhythmic tasks at about third grade and showed no improvement thereafter. The plateau in pitch ability was reached by the end of fourth grade, in general, with many children unable to reach the criterion on phrase-reproducing tasks at any age. Petzold concludes that age is a significant factor in the development of auditory perception. Mainwaring (35), in a study with children nine-and-one-half- to eleven-and-one-half-years-old, found that no consistent relationship exists between perception of pitch differences and perception of rhythmic patterns, and that age is an important factor in the development of cognitive abilities of this kind.

Less directly related are studies by Bond and Simpson in the area of gross motor performance and locomotor response to rhythmic stimuli. Bond (6) computed correlations between scores from the Seashore Test of Rhythm, given in the written form and with an apparatus which presented the rhythmic patterns in aural, visual, and tactile modes, and scores on measures of motor performance and motor learning ability. She found no significant relationship between rhythmic perception and motor performance or motor learning ability. Simpson (49) used an electrical device with two spring-mounted platforms connected to a kymograph for objectively recording subject's locomotor (foot movement) responses to various rhythmic stimuli. She found low correlation between scores on locomotor response and scores on the time discrimination and rhythm identification sections of the Kwalwasser-Dykema Music Tests. Both studies suggest that cognitive perception of rhythm is not significantly related to motor response to rhythm.

In summary, it appears that very few studies in musical concepts, as measured in the present study, have been undertaken. This absence of direct related studies is graphically illustrated in The Encyclopedia of Educational Research (45). In this comprehensive review of concept literature, no reference is made to studies of musical concepts.

CHAPTER III

DEVELOPMENT OF BATTERY OF MUSICAL CONCEPT MEASURES

Verbal Measure

The Verbal Measure of eighteen multiple-choice items was developed for use in a group or classroom situation. Containing six items for measuring each of the dimensions (i.e., the elements of pitch, duration, and loudness), the measure involved a comparison and discrimination of natural or music-related sounds recalled from prior experience. This task required a minimal ability to read and comprehend, as well as a background of experience with natural and musical sounds. In the construction of the measure, a rigorous attempt was made to limit the comparisons to sounds and the reading to words the subjects were likely to have encountered in their everyday existence.

The final form of the Verbal Measure evolved from several test-and-revise cycles between July of 1965 and August of 1966. During the summer of 1965 approximately one hundred verbal items were devised, proposed, and evaluated by members of the research team. After considerable discussion of content and form, the five-answer multiple-choice format was adopted. The use of "none of these" as an answer choice was excluded because of the possibility of confusing or frustrating subjects with its negative connotation. However, the answer "any of these" was included in some items because it seemed to allow subjects to demonstrate possession of the concepts.

University Area Trial, August 1965

Purpose. The purpose of this exploratory trial was (1) to try the verbal items with elementary school pupils, (2) to discover the length of time needed for administering the measure, and (3) to estimate the verbal difficulties subjects might encounter with the items.

Subjects. The subjects in this sample were twenty volunteers from communities near The Pennsylvania State University, about equally divided between children who would soon enter and those who had recently completed fourth grade.

Materials and Procedures. The measure used in this trial included seventeen multiple-choice items. Subjects were instructed to mark one answer for each item, choosing the best answer from five choices. The items included in this form appear as Numbers 1 to 17 in Verbal Measure, Form 1 (Appendix B). This group required approximately fifteen minutes to complete the measure.

Results. An analysis of results from this trial revealed a very high proportion of correct responses, scores ranging from nine to seventeen with a mean of 14.75 and a standard deviation of 2.35.

Discussion. The subjects gave little indication of difficulty in reading and comprehending the items. Because of the nature of the sample group, the results obtained in this trial were inconclusive and further trials with a more representative sample were projected. Six additional items were constructed to provide a pool of items adequate for choosing a balanced number of tested items measuring concepts of the three dimensions.

Estella Trial, September 1965

Purpose. The purpose of this trial was (1) to try the items in the Verbal Measure with a non-volunteer group of fourth-grade pupils and (2) to obtain data for statistical analysis.

Subjects. The subjects were the entire fourth grade of thirty-two pupils[1] in the Loyalsock Area Elementary School, Sullivan County School District, Estella, Pennsylvania. The area is rural, sparsely-populated, and situated a considerable distance from any urban center.

Materials and Procedures. The measure used in this trial was in two parts: Items 1 through 17 comprised one part, Items 18 through 23 a second part (Appendix B, Verbal Measure, Form 1). Because of interruptions in the school schedule, only the first part of the measure was administered during the first testing session, and the six additional items were completed in a later session the same day. Instructions were to mark the best answer for each item, marking only one answer for each. Subjects were instructed to request help during the testing if they experienced difficulty reading any words. These words were then read to them individually by an examiner.

Results. The mean score for the trial was 9.65, with a standard deviation of 3.86. The scores, divided into the two administration segments, showed a mean of 8.27, a standard deviation of 3.40, and a Kuder-Richardson reliability[2] of .67 for the first seventeen items; a mean of 1.38 and a standard deviation of 1.26 for the last six items. An item analysis (Appendix A, Table LI) for the measure showed item difficulty ranging from .00 to .90 for this administration, with two items answered correctly by less than 10 per cent of the subjects and two items answered correctly by more than 65 per cent of the subjects. The discrimination indices ranged from .00 to .82, with two items showing values less than .39.

---

[1] In this and subsequent trials, any discrepancies between the number of subjects tested and the "N's" appearing on various statistical tables are due to deletion of subjects for one of these reasons: (1) known hearing defects, (2) complete data not available, (3) scores or other data not usable for a specific analysis, (4) failure to observe specified test procedure.

[2] Robert L. Thorndike and Elizabeth Hagen, Measurement and Evaluation in Psychology and Education (second edition; New York: John Wiley and Sons, Inc., 1961), p. 181. The formula used here, and designated in subsequent computations as Kuder-Richardson reliability, is Kuder-Richardson Formula 21.

Discussion. As a result of this administration, the general format and methods of procedure were accepted for use with subsequent groups. Most of these subjects had no difficulty following directions for marking answers, and few indicated that they could not read or understand the items. With two exceptions, Items 4 and 21, the items discriminated well between the high and low scorers, and most items showed an acceptable level of difficulty.

The problem which seemed most in need of further investigation was the apparent inconsistency in performance of individual pupils on the two segments of the test. In order to ascertain whether the discontinuous testing procedure contributed to the inconsistent performance on the items, administration to another sample was planned.

Central Dauphin Trial, October 1965

Purpose. The purpose of this trial was (1) to test the twenty-three items of the Verbal Measure as a unit in a single, continuous session and (2) to obtain additional information on the difficulty and discriminating power of the individual items before revising the measure for the Pilot Study.

Subjects. The subjects were twenty-four pupils in one fourth-grade classroom in the Phillips School, Central Dauphin School District, Harrisburg, Pennsylvania. This district is part of a large, urban area.

Materials and Procedures. The measuring instrument for this administration was identical to that used in the Estella trial, but was administered in one session (Appendix B, Verbal Measure, Form 1). The administrator in this trial provided opportunity for pupils to ask questions concerning procedure.

Results. The mean score for this trial was 10.96, with a standard deviation of 4.02, and a Kuder-Richardson reliability of .67. For the first seventeen items a mean of 9.46, a standard deviation of 2.9, and a reliability of .53 were computed. The final six items showed a mean of 1.5, a standard deviation of 1.1, and a reliability of .09.

An item analysis (Appendix A, Table LII) showed three items answered correctly by less than 20 per cent and four items answered correctly by more than 80 per cent of subjects. Discrimination indices showed two items below .15, three items between .16 and .30, and all remaining items between .30 and .85.

An item analysis computed from the combined scores of the Central Dauphin and Estella trials (Appendix A, Table LIII) showed three items answered correctly by less than 20 per cent and two items answered correctly by more than 80 per cent of the subjects. Three items had discrimination indices below .20, three items were between .25 and .35, and all remaining items had discrimination indices from .40 to .79.

Discussion. A comparison of the scores of the two samples on the final six items of the Verbal Measure revealed a very small difference in the means of 1.5 and 1.38. This was indication that the items themselves, rather

than the testing conditions, were responsible for the difference in the scores on the two segments of the measure.

As a result of the item analysis, six items were discarded as non-discriminating or as inappropriate in difficulty for fourth-grade pupils. Items were revised or rewritten and one new item constructed to provide the needed quota of six items involving concepts of each dimension within the balanced measure of eighteen multiple-choice items. Answer choices for each item were checked to ascertain which had been functioning effectively as decoys. Alternatives not chosen by subjects were deleted and new choices substituted. The eighteen revised items were ordered randomly, a sample item included, and written instructions for marking answers developed and added, completing the preparation of the measure for the Pilot Study (Appendix B, Verbal Measure, Form 2).

Pilot Study, January 1966

Purpose. The purpose of this administration was (1) to examine the items and procedure for the administration of the measure using a random sample of fourth-grade pupils and (2) to obtain data for statistical analysis.

Subjects. The subjects were fifty-six pupils in two randomly-selected fourth-grade classrooms, one rural and one urban, in the Washington County, Maryland, schools.[3]

Materials and Procedures. The instrument used in this administration was Verbal Measure, Form 2, as described. The administrator read aloud the directions and the sample item. After pupils had checked an answer choice, they were told the correct answer. Pupils were then given an opportunity to ask questions relating to procedure. When all questions were answered, pupils continued with the eighteen items. If individual subjects indicated difficulty in reading words in any item, the words were read to them by a member of the research team. The total time used to complete the items was approximately twenty minutes, although some subjects finished in considerably less time.

Results. Means, standard deviations, and a Kuder-Richardson reliability estimate were computed for the complete measure and for the separate dimensions. Table I shows these results for the entire group participating in the classroom testing of the Written Measures (N = 56) and also for the randomly-selected sample who completed all four measures (N = 38).

An item analysis (Table II) showed item difficulty ranging from .26 to .89, with one item answered correctly by less than 30 per cent of subjects and two items answered correctly by more than 80 per cent of subjects. Discrimination indices ranged from .36 to .81, with only one item below .40.

---

[3]This district televises elementary music instruction from a central location to elementary schools in the system.

Discussion. On the basis of item analysis data, items were revised to increase or decrease item difficulty where this deviated substantially from .50. Because several of the answer choices appeared not to represent easily-identifiable sounds to some subjects, as indicated by the frequent selection of certain incorrect answers, two items were deleted. New items were devised to replace these; changes were made in the answer choices for Item 1, and non-functioning decoys were altered on several other items. All items were checked and rechecked for clarity of expression and exactness of meaning by members of the research team and by a group of experienced music educators, graduate students at The Pennsylvania State University, during the summer of 1966. The agreement of the research team and the music educators on the items was accepted as evidence of content validity.

TABLE I

MEANS, STANDARD DEVIATIONS, AND RELIABILITY ESTIMATES FOR THE VERBAL MEASURE PILOT STUDY

Dimension	Number of items	N = 56			N = 38		
		Mean	S.D.	r_{11}*	Mean	S.D.	r_{11}**
Pitch	6	3.67	1.77	.722	4.16	1.65	.639
Duration	6	3.82	1.31	.318	3.92	1.57	.534
Loudness	6	3.53	1.58	.495	3.97	1.42	.404
Total	18	10.89	4.15	.794	12.05	4.01	.817

*Kuder-Richardson reliability estimate
**Strata fixed reliability estimate from RELIB, a computer program at The Pennsylvania State University (Appendix E)

In preparing the final form of the Verbal Measure (Appendix C, Verbal Measure, Main Study), a standardized wording was used for similar items (e.g., Items 3, 7, 8, and 14) to minimize reading and comprehension difficulties. Punctuation was standardized, and critical words in several stems were underlined to increase the probability that they would be noticed by subjects. This form of the measure was accepted for use in the Main Study.

Listening Measure

The Listening Measure was developed to measure the subject's ability to identify changes in the dimensions of pitch, duration, and loudness within the multidimensional frame of reference of orchestral music. In the final form it consisted of eighteen short musical items, from four to twenty seconds in length, excerpted from standard orchestral literature.

TABLE II

VERBAL MEASURE ITEM ANALYSIS*
PILOT STUDY
N = 56

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.73	1.00	.89	.59
2	P	.33	.92	.65	.63
3	D	.20	.73	.46	.53
4	D	.13	.80	.45	.66
5	D	.53	.92	.75	.49
6	L	.20	.86	.54	.65
7	L	.20	.53	.36	.36
8	D	.40	.92	.69	.58
9	P	.53	1.00	.82	.71
10	D	.53	.92	.75	.49
11	L	.27	.92	.62	.67
12	L	.13	.92	.54	.76
13	P	.33	.73	.53	.40
14	L	.00	.67	.26	.79
15	P	.40	1.00	.77	.76
16	D	.40	1.00	.77	.76
17	P	.13	.92	.54	.76
18	P	.27	1.00	.71	.81

*Chung-Teh Fan, Item Analysis Table (Princeton, New Jersey: Educational Testing Service, 1952).

These items were of three types:

1. In the first group of six items, each contained a predominant change in one dimension, requiring a single judgment of the predominant change within the musical example, e.g., faster, lower, softer.

2. In the second group of six items, each contained predominant changes in two dimensions, requiring a judgment of two changes within the example, e.g., higher and louder, faster and softer, lower and slower.

3. In the third group of six items, each pair of musical excerpts differed predominantly in one dimension, requiring judgment of the predominant manner in which the second excerpt differed from the first.

In scoring the final form, items in Groups 1 and 3 received one point each; items in Group 2 received two points each.

In developing the Listening Measure, musical items (excerpts from larger works) were identified as definite examples of change in a single dimension or in two dimensions. These excerpts were analyzed and discussed as possible items, and decisions made to retain them as acceptable or to reject them as ambiguous or as too easy or too difficult. Twenty-four trial items were finally accepted by consensus of the research team. These items constituted the preliminary measure administered to trial groups at The Pennsylvania State University.

Exploratory Trial, August 1965

Purpose. The purpose of this exploratory trial was (1) to discover a usable format and procedure and (2) to obtain an objective indication of agreement on the predominant changes in the excerpts selected for trial.

Subjects. The sample consisted of twelve music educators in a graduate course in music education at The Pennsylvania State University (designated as Mu Ed 574 in Table LIV).

Materials and Procedures. The twenty-four items (musical excerpts) and one sample item, in taped form, were used for this trial. Answer sheets were in the form of numbered blanks on which subjects wrote free-response-type answers (Appendix B, Listening Measure, Form 1). Subjects were instructed to listen to the example and describe the musical change.

Results. In scoring this measure, each item received one point. Scores ranged from eleven to twenty-three, with a mean of 18.1 and a standard deviation of 3.77. The percentage of correct answers for twenty of these items appears in Table LIV (Appendix A).

Discussion. Responses indicated a general misunderstanding of the type of judgment desired. To reduce the procedural difficulties, more explicit instructions were developed and a multiple-choice answer format devised. A revised trial measure was developed, eliminating four suspect items which had elicited a variety of responses. The revised taped measure of

twenty items included an additional sample item and revised spoken directions, identical with those on the answer form.

University Groups Trials, August 1965

Purpose. The purpose of the trials was (1) to test the revised format and procedure and (2) to obtain evidence of agreement or disagreement on predominant musical changes in the excerpts from subjects who were not music educators.

Subjects. The two groups of subjects from The Pennsylvania State University were:

1. Eighteen elementary classroom teachers and prospective classroom teachers in an undergraduate course in methods for teaching elementary music (designated as Mu Ed 86 in Table LIV).

2. Twenty-eight high school students in the chorus of the Summer Band-Orchestra-Chorus School (designated as BOC Chorus in Table LIV).

Materials and Procedures. The revised instrument, including twenty items, two sample items, and directions for procedure, was used in these two trials, with a copy of the multiple-choice format answer sheet provided for each subject.

Results. The range of correct responses on the twenty items for Mu Ed 86 was from six to twenty, with a mean score of 15.7 and a standard deviation of 3.22; scores for BOC Chorus ranged from thirteen to twenty, with a mean of 17.43 and a standard deviation of 1.72. The percentage of each of the three trial groups (Mu Ed 574, Mu Ed 86, and BOC Chorus) correctly answering each item appears in Table LIV (Appendix A). A correlation coefficient (Pearson's *r*) of .804 was computed between the scores of Mu Ed 86 and BOC Chorus.

Discussion. A comparison of scores from Mu Ed 86 and BOC Chorus indicated that, although the percentage of the BOC Chorus answering correctly was generally higher, the overall performance of the two groups showed considerable agreement on the relative difficulty of the items. Examination of the percentages of all three groups revealed that a substantial majority of each group chose the correct answers. This agreement indicated that the predominant change in these items was discernible.

The discrepancy between the scores of the Mu Ed 574 sample and the other two groups was a matter of concern. Results seemed to indicate that, in addition to the possible effects of change in procedure and a different method of answering, the group of music educators may have been noticing aspects of the music differing from those apparent to the two less experienced groups.

Using the results of these first three trials as a guide to usable content and format, a revised set of items was prepared for a trial with elementary school pupils. Items showing a relatively low percentage of correct

answers were omitted, and new items judged acceptable by consensus of the research team were added to produce a total of twenty-four items. A decision was made to remove a stimulus variable by using only orchestral examples. To achieve homogeneity of medium, non-orchestral excerpts and orchestral examples containing solo passages were deleted, and all items subsequently chosen utilized the orchestral medium.

University Area Trial, August 1965

Purpose. The purpose of this exploratory trial was (1) to test items and format with elementary school subjects and (2) to identify procedural difficulties and non-usable items.

Subjects. The subjects in this sample were twenty volunteers from communities near The Pennsylvania State University, about equally divided between children who would soon enter and those who had recently completed fourth grade.

Materials and Procedures. The measure used in this trial consisted of twenty-four items. It included a tape containing twenty-four musical excerpts, spoken directions, sample excerpts, and spoken citation of item numbers, as well as answer sheets with directions for marking answers (as spoken on the tape) and five answer choices for each item (Appendix B, Listening Measure, Form 3). Subjects were instructed to listen to each item, choose the best answer, then listen again before marking the answer.

Results. The results of this trial revealed a high proportion of correct answers, scores ranging from eleven to twenty-two, with a mean of 17.4 and a standard deviation of 3.38. The results of an item analysis for this trial appear in Table LV (Appendix A). An examination of this data showed that only three of the twenty-four items were answered incorrectly by more than 50 per cent of the subjects.

Discussion. In evaluating the results of this trial, the research team considered a number of possible reasons for the high mean. Among these were:

1. The non-random technique used in recruiting subjects biased the sample.

2. The presence in the sample of a substantial proportion of children entering fifth grade inflated the mean.

3. The measure in this form was too easy for fourth-grade children.

To further investigate these possibilities, it was decided to obtain a more representative sample of fourth-grade children for another trial.

After careful review of the procedural difficulties encountered during the several early trials of the Listening Measure, a new tape was developed including revised directions and additional sample items. It was decided that the final form of the Listening Measure would consist of eighteen items, six requiring a judgment of a single change within each excerpt, six requiring

a judgment of two concurrent changes within each excerpt, and six requiring a judgment of a single difference between the two examples in a pair of excerpts. The first and third groups were to contain one item for each of six possible changes: higher, lower, louder, softer, faster, slower. For the second group, items with two changes, six of the possible answer combinations were selected randomly, within the requirement that resulting combinations provide an equal number of answers for each dimension (i.e., pitch, duration, loudness). In this third trial instrument, however, the full twenty-four items from the previous trial were included to provide a larger number of tested items for analysis and to provide a basis for comparison of scores from the two trials.

Estella Trial, September 1965

Purpose. The purpose of this trial was (1) to test the listening items with a non-volunteer group of fourth-grade children and (2) to obtain data for statistical analysis.

Subjects. The subjects were the entire fourth grade of thirty-two pupils in the Loyalsock Area Elementary School, Sullivan County School District, Estella, Pennsylvania.

Materials and Procedures. The measure used for this administration was the form revised after the preceding trial (Appendix B, Listening Measure, Form 3). The twenty-four items were unchanged, but additional sample items and directions were added. In the administration, the experimenter gave general directions in addition to the instructions on the tape, and assisted pupils with the sample items.

Results. The mean score for this trial was 9.50 correct responses, with a standard deviation of 4.84 and a Kuder-Richardson reliability estimate of .787. Results of an item analysis showed item difficulty ranging from .05 to .70 and discrimination indices ranging from -.27 to .93. With the exception of four items, all were .39 or above in discrimination (Appendix A, Table LVI).

Discussion. The mean score for this sample was substantially lower than the 17.4 obtained in the University Area sample. This may have been due in part to the non-select nature of the Estella group, which included all pupils in the fourth grade, to the difference in maturity between fourth and fifth graders, to the difference between the two groups in musical background and general cultural environment, or to a combination of these factors. A third trial was scheduled using an urban sample to further test the items.

Central Dauphin Trial, October 1965

Purpose. The purpose of this trial was (1) to test the items in the Listening Measure with a second non-volunteer group of fourth-grade subjects and (2) to obtain additional information on the difficulty and discriminating power of the individual items before revision of the measure for the Pilot Study.

Subjects. The subjects were twenty-four pupils in one fourth-grade classroom in the Phillips School, Central Dauphin School District, Harrisburg, Pennsylvania.

Materials and Procedures. The instrument and procedures for this trial were identical to those used in the Estella trial.

Results. The mean score for this trial was 11.96, with a standard deviation of 4.46 and a Kuder-Richardson reliability estimate of .73. Scores from this sample and the Estella trial were combined to provide a larger sample. The mean score for the combined sample was 10.59, with a standard deviation of 4.79 and a Kuder-Richardson reliability estimate of .77 (Appendix A, Table LVII). An item analysis was computed for the combined data (Appendix A, Table LVIII).

Discussion. The means from the Estella and Central Dauphin trials indicated that the Listening Measure, as a whole, was of appropriate difficulty for fourth-grade children. The results of the item analysis were used in selecting items for use in the Pilot Study; those items showing a low discrimination index or a high difficulty level were not included in the revised form of this measure.

Using the best items from this trial measure, as determined by the item analysis, and adding new ones as necessary, a balanced measure of eighteen items was constructed. This measure contained six single excerpts requiring single answers, six single excerpts requiring double answers, and six pairs of excerpts for comparison requiring single answers. Because of the six double-answer items, the total possible score on the measure was twenty-four. The measure was balanced with eight items involving pitch concepts, eight duration concepts, and eight loudness concepts (Appendix B, Listening Measure, Form 4).

Pilot Study, January 1966

Purpose. The purpose of this administration was (1) to examine the items and procedure for the administration of the measure using a random sample of fourth-grade pupils and (2) to obtain data for statistical analysis.

Subjects. The subjects were fifty-six pupils in two randomly-selected fourth-grade classrooms, one rural and one urban, in the Washington County, Maryland, schools.

Materials and Procedures. The measure used in this trial consisted of eighteen items. It included a tape containing eighteen musical excerpts, spoken directions, sample excerpts, and spoken citation of item numbers, as well as answer sheets in five-alternative multiple-choice form. (The musical source of items, the directions for administration, and an answer sheet appear in Appendix B, Listening Measure, Form 4.) Subjects were encouraged to ask questions concerning the procedure after each set of instructions and sample items was completed. The testing time was approximately twenty-five minutes, including time for directions and sample items.

Results. Means, standard deviations, and a Kuder-Richardson reliability estimate were computed for the complete measure and for the separate dimensions. Table III shows these results for the entire group participating in the classroom testing of the Written Measures (N = 56) and for the randomly-selected sample completing all four measures (N = 38).

TABLE III

MEANS, STANDARD DEVIATIONS, AND RELIABILITY ESTIMATES FOR THE LISTENING MEASURE PILOT STUDY

Dimension	Number of items	N = 56			N = 38		
		Mean	S.D.	r_{11}*	Mean	S.D.	r_{11}**
Pitch	8	4.32	2.23	.69	5.05	2.32	.76
Duration	8	5.04	1.92	.56	5.47	1.75	.60
Loudness	8	5.40	1.87	.57	6.13	1.14	.34

*Kuder-Richardson reliability estimate

**Strata fixed reliability estimate from RELIB, a computer program at The Pennsylvania State University (Appendix E)

An item analysis (Table IV) showed item difficulty ranging from .26 to .85, with one item answered correctly by less than 30 per cent of subjects and one answered correctly by more than 80 per cent of subjects. Discrimination indices ranged from .21 to .89.

Discussion. In preparing the Listening Measure for the Main Study, items were examined with reference to the results of the Pilot Study item analysis. It was decided (1) to delete any items with a discrimination index less than .30 and (2) to examine items with a proportion of correct answers smaller than .30 or greater than .75 for possible replacement.

After revisions were made, a trial instrument was administered to seventeen music educators during the summer of 1966 for preliminary validation (Appendix B, Answer Sheet, Preliminary Validation). With the results of this procedure and the Pilot Study as a guide, three of the items were replaced and the directions were revised. It was also decided that more time was needed between items, and that ten seconds was sufficient to give subjects time to read answer choices, decide, and mark the correct answer. (Revised directions appear in Appendix C, Listening Measure Administration, Main Study.) The spoken direction, "Mark your answer for number ___", was included on the tape after the second playing of each item.

The final form of the Listening Measure (Appendix C, Listening Measure, Main Study) was prepared with the cooperation of the Division of Instructional Services of The Pennsylvania State University.

TABLE IV

LISTENING MEASURE ITEM ANALYSIS*
PILOT STUDY
N = 56

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.46	.86	.67	.45
2	P	.066	1.00	.59	.89
3	P	.46	1.00	.79	.74
4	L	.066	.67	.34	.65
5	D	.00	.67	.26	.79
6	D	.60	1.00	.85	.67
7	D + L	.20	.92	.58	.71
8	P + L	.40	1.00	.77	.76
9	D + L	.40	.92	.69	.58
10	P + D	.20	.86	.54	.65
11	P + D	.13	.86	.49	.71
12	P + L	.20	.67	.43	.48
13	D	.60	.86	.74	.32
14	P	.20	.92	.58	.71
15	L	.33	.86	.61	.55
16	P	.13	.46	.28	.39
17	L	.46	1.00	.79	.74
18	D	.33	.53	.43	.21

*Fan table

Validation Trial, August 1966

Purpose. The purpose of this trial was (1) to test the revised answer-sheet format and the procedure for administration and (2) to ascertain whether the Listening Measure items selected for the Main Study would elicit correct judgments of a musical change under test conditions.

Subjects. The subjects for this trial were forty-five high school students participating in the Summer Music Clinic at The Pennsylvania State University.

Materials and Procedures. The Listening Measure in the form to be used in the Main Study, consisting of five sample items and eighteen test items, was used for this trial. Answer sheets and administration procedures followed the revised forms (Appendix C, Listening Measure, Main Study).

Results. Table V shows the percentage of students correctly answering the items of the measure.

TABLE V

PERCENTAGE OF CORRECT ANSWERS TO LISTENING ITEMS
HIGH SCHOOL STUDENTS
N = 45

Item number	Per cent correct	Item number	Per cent correct	Item number	Per cent correct
1	100.0	7	95.6	13	93.3
2	95.6	8	100.0	14	97.8
3	97.8	9	100.0	15	91.1
4	82.2	10	100.0	16	77.8
5	100.0	11	100.0	17	97.8
6	100.0	12	84.4	18	88.9

Discussion. A high proportion of correct answers was achieved by this group. The results indicated that the items did change predominantly in the specified dimensions, and the listeners could perceive these changes under conditons similar to classroom testing. This was accepted as evidence of content validity. Instructions, content, and format of answer sheet worked satisfactorily with the group. A decision was made to utilize the Listening Measure in this form (items, answer sheets, directions, and procedures) for the Main Study.

Manipulative Measure

The Manipulative Measure was developed for use with individual subjects and provided a mode of demonstrating understanding of the musical concepts of pitch, duration, and loudness through the manipulation of simple music-making devices (e.g., triangle, resonator bells, finger cymbal). Items were devised to measure these concepts, and promising items were selected by consensus of the research team.

After several trials and revisions, the final form of the Manipulative Measure contained eighteen items, six items involving each of the dimensions of pitch, duration, and loudness. Items were designed so that the manual dexterity required for performing each test item was minimal.

University Area Trial, August 1965

Purpose. The purpose of this exploratory trial was (1) to try the items devised and estimate their effectiveness and (2) to try the procedure for the administration of the measure.

Subjects. The subjects were eleven volunteer children, from communities near The Pennsylvania State University, who would soon enter or had recently completed fourth grade.

Materials and Procedures. The measure used in this trial consisted of fifteen items. Subjects were encouraged to experiment and become familiar with the music-making devices. Each task was explained by the experimenter, and the subject attempted to perform the required task (Appendix B, Manipulative Measure, Form 1). One experimenter administered the measure to individual subjects and judged the responses right or wrong for a possible total of fifteen points; other experimenters observed the administration of the measure.

Results. The results of this trial showed a high proportion of correct answers, with scores ranging from nine to fifteen, a mean of 13.27, and a standard deviation of 1.77. Items were examined and several were discarded because of failure to elicit responses from the subjects.

Discussion. New items were devised, examined, and incorporated into the measure, which was then balanced with six items measuring pitch, six duration, and six loudness concepts, a total of eighteen items (Appendix B, Manipulative Measure, Form 2). The procedure described above seemed to be effective in permitting subjects to demonstrate possession of the concepts and was accepted for subsequent use. Since the subjects in the first trial had been volunteers, further trials with more representative subjects were planned.

Estella Trial, September 1965

Purpose. The purpose of this trial was (1) to test the items with a non-volunteer group of fourth-grade children and (2) to provide data for statistical analysis.

Subjects. The subjects were the entire fourth-grade of thirty-two pupils in the Loyalsock Area Elementary School, Sullivan County School District, Estella, Pennsylvania.

Materials and Procedures. The measure used in this trial consisted of eighteen items (Appendix B, Manipulative Measure, Form 2). Time for administration of the measure varied from ten to fifteen minutes. The items were judged right or wrong, with two experimenters alternately judging and observing.

Results. Results of this trial showed a mean of 12.58, a standard deviation of 2.92, and a Kuder-Richardson reliability estimate of .574. An item analysis (Appendix A, Table LIX) showed that loudness items had discrimination indices ranging from .00 to .72; pitch items had indices ranging from .63 to .93; duration items had indices ranging from .00 to .82; and all items showed item difficulty ranging from .24 to 1.00.

Discussion. This trial provided the research team with an evaluation of the items in the measure. It was decided to delete Item 7 and decrease from six to four the number of resonator bells used in Items 13 through 18; these items had proved too difficult. New items were developed and incorporated into a revised measure (Appendix B, Manipulative Measure, Form 3).

Bellefonte Area Trial, October 1965

Purpose. The purpose of this trial was to test the effectiveness of the items prior to their incorporation into the measure for the Pilot Study.

Subjects. The subjects were twenty randomly-selected pupils from a fourth-grade classroom in the Bellefonte Elementary School, Bellefonte Area School District, Bellefonte, Pennsylvania.

Materials and Procedures. The measure used in this trial consisted of eighteen items, six measuring pitch, six duration, and six loudness concepts (Appendix B, Manipulative Measure, Form 3). Time for administration of the measure varied from ten to fifteen minutes. The experimenter judged the responses right or wrong.

Results. The results of this trial showed a mean of 11.3, a standard deviation of 2.44, and a Kuder-Richardson reliability estimate of .31. Item analysis (Appendix A, Table LX) indicated that most of the items discriminated with this group.

Discussion. Since the item analysis indicated that most of the items discriminated with this group, the Manipulative Measure, Form 3, was accepted for the Pilot Study.

Pilot Study, January 1966

Purpose. The purpose of the Pilot Study was (1) to examine the items and the procedure for the administration of the measure using a random sample of fourth-grade pupils and (2) to obtain data for statistical analysis.

Subjects. The subjects were thirty-eight randomly-selected pupils from two randomly-selected fourth-grade classrooms in two schools, one rural and one urban, in the Washington County, Maryland, School District.

Materials and Procedures. The measure consisted of eighteen items with six items measuring pitch, six duration, and six loudness concepts (Appendix B, Manipulative Measure, Form 3). Subjects were encouraged to experiment and become familiar with the music-making devices. After each task was explained by the experimenter, the subject attempted to perform the required task. Items were judged right or wrong for a possible total of eighteen points. Time for administration of the measure varied from ten to fifteen minutes.

Results. The means, standard deviations, and Kuder-Richardson reliability estimates of the Pilot Study are shown in Table VI.

TABLE VI

MEANS, STANDARD DEVIATIONS, AND RELIABILITY ESTIMATES FOR THE MANIPULATIVE MEASURE PILOT STUDY N = 38

Dimension	Number of items	Mean	Standard deviation	Reliability estimate*
Pitch	6	3.08	1.95	.872
Duration	6	4.90	1.29	.535
Loudness	6	5.74	.50	.03
Total	18	13.71	3.09	.696

*Strata fixed reliability estimate from RELIB, a computer program at The Pennsylvania State University (Appendix E)

Item analysis (Table VII) showed the pitch items had discrimination indices ranging from .62 to .88, duration items had indices ranging from .39

TABLE VII

MANIPULATIVE MEASURE ITEM ANALYSIS*
PILOT STUDY
N = 38

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	D	.40	1.00	.77	.76
2	D	.40	1.00	.77	.76
3	L	1.00	1.00	1.00	.00
4	D	.90	1.00	.95	.39
5	D	.90	1.00	.95	.39
6	P	.20	1.00	.67	.84
7	P	.10	1.00	.61	.88
8	P	.30	1.00	.72	.80
9	P	.10	1.00	.61	.88
10	L	1.00	1.00	1.00	.00
11	L	.80	1.00	.92	.53
12	P	.20	1.00	.67	.84
13	L	.80	1.00	.92	.53
14	D	.60	1.00	.85	.67
15	L	.90	1.00	.95	.39
16	P	.10	.70	.38	.62
17	L	.90	1.00	.95	.39
18	D	.20	1.00	.67	.84

*Fan table

to .84, loudness items had indices ranging from .00 to .53, and all items showed item difficulty ranging from .26 to .85.

Discussion. The procedure described was satisfactory in permitting subjects to demonstrate possession of the concepts. The item analysis indicated that the loudness items and two of the duration items had unacceptable difficulty and discrimination indices. Procedures for the revision of these items included:

1. Changing the directions for Item 11 from "Which one is loudest?" to "Play from the loudest to the softest."

2. Changing the directions for Item 13 from "Find the softest position." to "Find the softest, medium, and loudest positions."

3. Changing the directions for Item 15 from "Which set is loudest?" to "Play in order from softest to loudest."

4. Changing the directions for Item 17 from "Which is softest?" to "Play in order from loudest to softest."

5. Re-examining the chord organ for a possible loudness item.

6. Re-examining all items to confirm the presence of at least two other dimensions to function as decoys for the dimension being measured.

Items were re-ordered and this revised measure (Appendix B, Manipulative Measure, Form 4) was submitted to a field trial.

Matternville Trial, May 1966

Purpose. The purpose of this trial was to examine the revised items.

Subjects. The subjects were eleven randomly-selected fourth-grade pupils from a classroom in the Matternville Elementary School, Matternville, Pennsylvania.

Materials and Procedures. The measure used in this trial contained eighteen items with six items involving concepts of pitch, six duration, and six loudness (Appendix B, Manipulative Measure, Form 4). The responses to the items were judged either right or wrong for a possible total of eighteen points. Subjects were encouraged to experiment and become familiar with the music-making devices. After each task was explained by the experimenter, the subject attempted to perform the required task. Time for administration of the measure varied from ten to fifteen minutes.

Results. The means, standard deviations, and Kuder-Richardson reliability estimates for the total measure and the dimensions of pitch, duration, and loudness are shown in Table VIII.

Discussion. An examination of the results indicated that the loudness items needed further revision. Items 7, 11, and 13 were deleted from

the measure, and the consideration of new loudness items continued. The Dynalevel, an electronic apparatus measuring discrete variations in the loudness of sounds, was used to estimate the loudness levels of various music-making devices. On this basis, five items were tentatively accepted for possible incorporation into the measure.

TABLE VIII

MEANS, STANDARD DEVIATIONS, AND RELIABILITY ESTIMATES FOR THE MANIPULATIVE MEASURE MATTERNVILLE TRIAL N = 11

Dimension	Number of items	Mean	Standard deviation	Reliability estimate*
Pitch	6	2.91	1.31	.334
Duration	6	4.36	1.61	.687
Loudness	6	5.36	.64	.193
Total	18	12.64	2.71	.649

*Kuder-Richardson reliability

Validation Trials, August 1966

The five new loudness items were submitted for validation to eleven faculty members and graduate students in music education (a jury of experts) at The Pennsylvania State University. The criterion for acceptance was agreement of at least nine of the eleven judges. As a result of this procedure, three items were accepted and incorporated into the measure.

The revised measure was then submitted to a second group of experts for validation. Results of the validation showed there was unanimous agreement by the experts on each item; this was accepted as content validity. The measure was accepted for the Main Study (Appendix C, Manipulative Measure, Main Study).

Overt Measure

The Overt Measure was developed to provide individual subjects with a mode of demonstrating understanding of the musical concepts of pitch, duration and loudness by responding with bodily movements to selected excerpts from orchestral literature. Musical examples were identified as items containing a predominant change in the musical dimensions being measured. These excerpts were analyzed and chosen by consensus of the research team and incorporated into the measure.

In the final form the Overt Measure consisted of nine musical excerpts ranging from thirteen to twenty-five seconds in length. Subjects were judged for any change of movement in response to the predominant change in the music and for an oral response to the change in the music.

University Area Trial, August 1965

Purpose. The purpose of this exploratory trial was (1) to examine several possible procedures for the administration of the measure and (2) to determine the difficulty and effectiveness of the items.

Subjects. Subjects were eleven volunteer children from the area near The Pennsylvania State University who would soon enter or had recently completed fourth grade.

Materials and Procedures. The measure contained seventeen excerpts (Appendix B, Overt Measure, Form 1). Various prccedures and different experimenters were used to elicit responses from the subjects. Among the procedures were:

1. The subject listened to the first playing of the item and moved to it on the second playing. Although this basic method was used for all subjects, some subjects heard the excerpt an additional time because of failure to respond on the second playing.

2. The subject observed while the experimenter demonstrated possible movements, showing a change where the music changed. The subject was then encouraged to move to the music.

3. Beginning with the fourth subject, the first six items were omitted and the measure initiated with Item 7, which was more rhythmic and had an impelling effect on the movements of the subjects.

Movement was observed to determine (1) the presence of change in movement with the change in the music, and (2) the type of movement exhibited by the subject, e.g. smooth to jerky, fast to slow, etc.

Results. Due to the exploratory nature of this trial, in which various procedures were employed, statistical analyses were not appropriate.

Discussion. Items 1 through 6 were deleted because these items did not elicit movement from the subjects. A revision of the measure was developed using Items 7 through 17 (Appendix B, Overt Measure, Form 2). Subjects were now to be judged on changing movement when the music changed (one point) and on the appropriateness of movement to the change in the music (one point). The procedure in which the subject listened to the excerpt on the first playing and moved to the excerpt on the second playing was adopted for subsequent use.

Estella Trial, September 1965

Purpose. The purpose of this trial was to examine the difficulty and effectiveness of the items with fourth-grade pupils.

Subjects. The subjects were an entire fourth grade of thirty-two pupils in the Loyalsock Elementary School, Sullivan County School District, Estella, Pennsylvania.

Materials and Procedures. The measure used in this trial consisted of eleven items, each item receiving two points for a possible total of twenty-two points (Appendix B, Overt Measure, Form 2). Subjects were judged on any change of movement to the predominant change in the music (one point) and on appropriateness of movement to this change (one point). Subjects were instructed to listen to the music on the first playing and move to the music on the second playing.

Results. The results showed a mean of 13.0, a standard deviation of 2.60, and a Kuder-Richardson reliability of .217. An item analysis (Appendix A, Table LXI) showed that all but one of the items had discrimination indices ranging from .63 to .93 and difficulty indices ranging from .37 to .75.

Discussion. Although the item analysis was favorable, it was not certain whether the response changes resulted from a change in the style of music or from a change in pitch, duration, or loudness. New items were identified and selected with particular attention to the problem of a single change in the music. The number of items in this revised measure was limited to nine because seventeen items had proved too fatiguing and time consuming. The measure (Appendix B, Overt Measure, Form 3) was balanced with three items measuring pitch, three duration, and three loudness concepts.

A classroom orientation was to be included prior to the administration of the measure to acquaint the subjects with possible physical responses. In this orientation a sample item would be played and verbal responses elicited from the subjects as to possible movement.

Bellefonte Trial, October 1965

Purpose. The purpose of this trial was to examine the judging of appropriateness of the overt response.

Subjects. The subjects were twenty randomly-selected pupils from a fourth-grade classroom in the Bellefonte Elementary School, Bellefonte Area School District, Bellefonte, Pennsylvania.

Materials and Procedures. The measure used in this trial contained nine items, each item receiving two points for a possible total of eighteen. Subjects were judged on any change of movement in response to the predominant change in the music (one point) and on appropriateness of movement to this change (one point). Subjects were instructed to listen to the music on the

first playing and move to the music on the second playing. Three judges rated certain subjects; two judges rated all twenty.

Results. The test for significant differences among judges, computed by an analysis of variance, yielded an F ratio of 1.19 (not significant at the .05 level) for the two judges, and an F of 3.58 (significant at the .05 level) for the three judges.

Discussion. Since the lack of agreement on the appropriateness of change contributed to the variability among judges, the judgment of this response was deleted. It was proposed that an oral response be substituted for the judgment of appropriateness. In this response the subject would be asked to indicate verbally what change occurred in the music. The six possible choices (higher, lower, louder, softer, faster, slower) would appear on a card shown to the subject. The proposed procedure for the administration was:

1. Listen to the excerpt.

2. Move to the music, changing when the music changes.

3. Listen again, refer to the card if necessary, and tell what change occurred.

Items 4, 6, and 7 were deleted from the measure because changes in these excerpts occurred gradually rather than at a definite point. Many excerpts in which the music had a definite point of change were examined and three new items accepted. A trial with another sample was planned with this revised measure (Appendix B, Overt Measure, Form 4).

Boalsburg Trial, November 1965

Purpose. The purpose of this trial was to examine the effectiveness of the oral response in relation to the overt response.

Subjects. The subjects were eight randomly-selected fourth-grade pupils from a classroom in the Boalsburg Elementary School, Boalsburg, Pennsylvania.

Materials and Procedures. The measure used in this trial contained nine items with each item receiving two points for a possible total of eighteen points. The subjects were judged on any change of movement to the predominant change in the music (one point) and for the correct oral response (one point). Subjects were instructed to (1) listen to the excerpt, (2) move to the music, changing when the music changes, and (3) listen again, refer to the card if necessary, and tell what change occurred.

Results. The results showed a mean of 5.4 and a standard deviation of .86 for changing when the music changed, and a mean of 4.25 and a standard deviation of 2.44 for the oral response.

Discussion. The results of this trial indicated that the oral response worked satisfactorily with this group. The oral response was then

incorporated into the measure and the revised measure accepted for the Pilot Study (Appendix B, Overt Measure, Form 4).

Pilot Study, January 1966

Purpose. The purpose of this administration was (1) to examine the items and the procedure for the administration of the measure using a random sample of fourth-grade pupils and (2) to obtain data for statistical analysis.

Subjects. The subjects were thirty-eight randomly-selected pupils from two randomly-selected fourth-grade classrooms in two schools, one rural and one urban, in the Washington County, Maryland, School District.

Materials and Procedures. The measure contained nine items, each item receiving two points for a possible total of eighteen points. The subjects were judged on any change of movement to the predominant change in the music (one point) and for the correct oral response (one point). Subjects were instructed to (1) listen to the excerpt, (2) move to the music, changing when the music changes, (3) listen again, refer to the card if necessary, and tell what change occurred. A classroom orientation to the measure was given at the completion of the group measures. In this procedure, pupils listened to a recorded example and suggested possible physical movements and the correct oral response.

Results. The means, standard deviations, and reliability estimates for the total measure and the dimensions of pitch, duration, and loudness are shown in Table IX.

TABLE IX

MEANS, STANDARD DEVIATIONS, AND RELIABILITY ESTIMATES FOR THE OVERT MEASURE PILOT STUDY N = 38

Dimension	Number of items	Mean	Standard deviation	Reliabiltiy estimate*
Pitch	6	1.90	1.50	.485
Duration	6	4.37	1.62	.644
Loudness	6	3.76	1.24	.080
Total	18	10.03	3.61	.479

*Strata fixed reliability estimate from RELIB, a computer program at The Pennsylvania State University (Appendix E)

TABLE X

OVERT MEASURE ITEM ANALYSIS*
PILOT STUDY
N = 38

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.35	.95	.69	.67
2	D	.60	1.00	.85	.67
3	D	.30	.90	.62	.62
4	L	.40	.75	.58	.36
5	D	.30	1.00	.72	.80
6	P	.05	.40	.20	.51
7	P	.05	.70	.33	.70
8	L	.45	.85	.66	.44
9	P	.20	.60	.39	.42

*Fan table

An item analysis (Table X) showed that the discrimination indices ranged from .36 to .80 and difficulty indices ranged from .20 to .85. Kuder-Richardson indices of agreement between the overt response and the oral response for the dimensions measured were: pitch, -.03; duration, .54; and loudness, .05.

Discussion. Although indices of agreement indicated that there was no significant relationship between the physical response and the oral response for the dimensions of pitch and loudness, the procedure using both overt and oral responses worked satisfactorily with these subjects. After an examination of the item analysis, Item 6 was deleted; a new item was developed and incorporated into the measure. A validation of the revised measure was planned.

Validation Trial, August 1966

The musical excerpts in the Overt Measure were submitted to seventeen graduate students in music education at The Pennsylvania State University for validation by experts. The results of this validation, Table XI, show that a high proportion of the experts agreed on the musical changes in the excerpts and the measure was accepted as having content validity.

TABLE XI

RESULTS OF VALIDATING TRIAL
OVERT MEASURE
N = 17

Item number	Correct response	Number responding correctly
1	softer	16
2	faster	17
3	slower	16
4	louder	17
5	slower	17
6	higher	14
7	lower	17
8	louder	17
9	higher	14

The final form of the Overt Measure, containing nine excerpts ranging from thirteen to twenty-five seconds in length (Appendix C, Overt Measure, Main Study), was prepared with the assistance of the Division of Instructional Services of The Pennsylvania State University.

Written and Nonwritten Measures

It was decided in December of 1965, after extensive consideration and consultation, to derive six scores from the Battery of Musical Concept Measures. The plan was to combine the pitch, duration, and loudness scores from the Verbal and Listening Measures by using stanines to produce derived scores for Written Pitch, Written Duration, and Written Loudness. The pitch, duration, and loudness scores from the Manipulative and Overt Measures were to be combined by using stanines to produce Nonwritten Pitch, Nonwritten Duration, and Nonwritten Loudness. Written scores would have a possible total of fourteen points and nonwritten scores a possible total of twelve points (Table XII).

Data from the Pilot Study were analyzed in terms of the written and nonwritten derived scores. Table XIII shows the means and standard deviations for the raw scores derived from the separate measures and the written and nonwritten derived scores.

Before combining the verbal and listening scores into a written score and combining the manipulative and overt scores into a nonwritten score, it was necessary to ascertain whether the means and standard deviations from the separate measures were equal for each dimension. The scores in Table XIII show that the means and standard deviations of the scores for the different dimensions, as derived from the measures of the battery, were not equal (e.g., the means and standard deviations for verbal pitch are not equal to those for listening pitch). Therefore, in order to have comparable scores for obtaining the written and nonwritten scores, it was decided to convert the derived scores from the separate measures for each dimension into a standard score. The stanine was the type of standard score selected, chosen to avoid a false aura of accuracy.

Stratified reliability estimates were obtained from the computer program RELIB, which gives a complete analysis of variance reliability (Appendix E). Table XIV shows the strata fixed reliability estimate, strata reliabilities, ratio of subjects by strata mean square to residual mean square, and correlations between strata for the Pilot Study sample as derived from this computer program.

The reliability estimates of all the measures, as indicated in Table XIV, are good except for Nonwritten Loudness, which shows low reliabilities for the Manipulative and Overt Measures. Strata exist for all written and nonwritten derived scores for the dimensions except Written Duration, which may indicate listening and verbal measurements of duration are not contributing different information.

The computer program COREL (Appendix E), used with the thirty-four students for whom complete data were available, produced the matrix in Table XV.

Table XV presents reliability and validity estimates. The underlined numbers are validity estimates, all significantly different from zero. Also included in this table are correlations of the written and nonwritten derived scores with IQ. Each of the written and nonwritten derived scores for the dimensions of pitch, duration, and loudness, except Nonwritten Loudness, correlates about .50 with IQ. Written Pitch, Written Loudness, and

Nonwritten Loudness showed correlations with IQ lower than their validity coefficients.

The Written and Nonwritten Measures generally showed a favorable analysis, which warranted use of this design in the Main Study.

DESIGN AND DEVELOPMENT OF THE BATTERY OF
MUSICAL CONCEPT MEASURES

	MEASURES			
	Written Total possible score for each dimension = 14		Nonwritten Total possible score for each dimension = 12	
Dimension	Verbal	Listening	Manipulative	Overt
Pitch	6	Single 2 Double 4 Comparison 2	6	Overt 3 Oral 3
Duration	6	Single 2 Double 4 Comparison 2	6	Overt 3 Oral 3
Loudness	6	Single 2 Double 4 Comparison 2	6	Overt 3 Oral 3
Total number of items	18	18 (Double items test two dimensions.)	18	9 (Each item requires two responses.)

TABLE XIII

MEANS AND STANDARD DEVIATIONS OF DERIVED SCORES,
WRITTEN, NONWRITTEN, AND SEPARATE MEASURES
PILOT STUDY
N = 38

Dimension		Verbal Measure (6 items)	Listening Measure (8 items)	Written Measures (14 items)	Manipulative Measure (6 items)	Overt Measure (6 items)	Nonwritten Measures (12 items)
Pitch	Mean	4.16	5.05	9.21	3.08	1.90	4.97
	S.D.	1.65	2.32	3.55	1.95	1.50	2.86
Duration	Mean	3.92	5.47	9.26	4.90	4.37	9.26
	S.D.	1.57	1.75	2.36	1.29	1.62	2.36
Loudness	Mean	3.97	6.13	10.11	5.74	3.76	9.50
	S.D.	1.42	1.14	2.24	0.50	1.24	1.43

TABLE XIV

STRATA FIXED RELIABILITY, STRATA RELIABILITY, RATIO OF SUBJECTS BY STRATA MEAN SQUARE TO RESIDUAL MEAN SQUARE, AND CORRELATIONS BETWEEN STRATA FOR THE WRITTEN AND NONWRITTEN DERIVED SCORES
PILOT STUDY
N = 38

Derived Score	Strata fixed Reliability	Strata Reliabilities		Ratio of Ss by Strata M.S. to Residual M.S.	Correlation Between Strata
		Verbal	Listening		
Written Pitch	.820	.661	.761	1.584	.499
Written Duration	.750	.597	.595	0.982	.616
Written Loudness	.594	.485	.340	1.098	.431
		Manipulative	Overt		
Nonwritten Pitch	.786	.762	.605	2.112	.391
Nonwritten Duration	.749	.596	.704	1.973	.351
Nonwritten Loudness	.284	-.007	.271	1.200	.106

TABLE XV

SYMMETRIC CORRELATION MATRIX
PILOT STUDY
N = 34

Measure	Written Pitch	Written Duration	Written Loudness	Nonwritten Pitch	Nonwritten Duration	Nonwritten Loudness
Written Duration	.741					
Written Loudness	.725	.687				
Nonwritten Pitch	.621	.657	.552			
Nonwritten Duration	.633	.403	.483	.623		
Nonwritten Loudness	.526	.280	.508	.505	.555	
IQ	.549	.544	.491	.500	.470	.263

CHAPTER IV

MAIN STUDY

Purpose

The purpose of the Main Study administration was to obtain data for statistical analyses including (1) reliability estimates for the four measures and for the derived scores of Written and Nonwritten Pitch, Duration, and Loudness, (2) item analysis of the several measures, (3) correlation coefficients and partial correlations between measures, derived scores, and scores from standard intelligence and reading achievement tests, and (4) construction of stanine conversion tables for the derived scores.

Subjects

Twelve Pennsylvania school districts, situated in a geographical area of approximately 23,000 square miles, were invited to participate in the Main Study. All twelve school districts, which represented various socioeconomic levels, agreed to participate. Administrators and music supervisors from the school districts were invited to the University Park Campus in May, 1966, to meet with the research team prior to the administration of the measures. They were briefed on the purposes and procedures of the project. Following this conference, information sheets and questionnaires were sent to the school districts (Appendix D).

On the basis of information received, one or two fourth-grade classrooms were randomly selected in each school district from buildings with adequate testing facilities, making a total of sixteen classrooms. Names of schools participating in the Main Study and the number of subjects from each school appear in Table XVI.

The group measures (Verbal and Listening) were administered to the entire classroom, following which fifteen or more pupils were randomly selected from the classroom sample. The individual measures (Overt and Manipulative) were administered to these pupils in the order of their selection. In most instances fewer than fifteen pupils were tested individually because of the time limitations of the school schedule.

Materials and Procedures

The Battery of Musical Concept Measures was administered in two parts as follows: (1) Verbal and Listening Measures were given to entire classroom groups and (2) Manipulative and Overt Measures were given individually to randomly-selected subjects in separate rooms. A general orientation

TABLE XVI

PENNSYLVANIA SCHOOL DISTRICTS
INCLUDED IN SAMPLE
MAIN STUDY

School district	Address	Building	Number of students	
			Total sample	Random sample
City of Allentown	Allentown	Muhlenberg	31	14
		Lehigh Parkway	25	13
Altoona	Altoona	Curtin	26	13
		Penn-Lincoln	29	15
Bellefonte Area	Bellefonte	Pleasant Gap	29	12
Central Cambria Joint	Ebensburg	Ebensburg Borough	23	16
Central Dauphin	Harrisburg	E. H. Phillips	24	13
Coatesville Area	Coatesville	Craig Ridgway	26	12
Greenwood	Millerstown	Liverpool	27	14
Lewistown-Granville-Rothrock Joint	Lewistown	Strodes Mills	37	12
State College Area	State College	Easterly Parkway	24	12
		Boalsburg Pike	18	17
Sullivan County	Dushore	Sonestown	29	14
West Shore	New Cumberland	Hillside	23	12
Williamsport Area	Williamsport	Hepburn-Lycoming	30	14
		Transeau	28	11

to the procedure and personnel preceded the group testing session. After being introduced to the class, the research team established a relaxed but cooperative atmosphere in the testing situation, telling subjects that the information was for research purposes and would not affect their grades in music. The administrator also explained that the measures were an attempt to find out what children knew and thought about music, and what they heard when listening to music. Children were supplied with pencils and erasers, and other physical arrangements were checked before beginning the Verbal Measure.

The Verbal Measure consisted of eighteen multiple-choice items plus one sample item. The form included printed directions for marking answers (Appendix C). The administrator read the printed directions aloud and helped children complete the sample item. Children were then to ask questions if they did not understand the procedure, and they were instructed to raise questions if difficulty was encountered reading test items.[1] When all preliminary questions were answered, the group proceeded with the eighteen items. When a child indicated difficulty reading words in any item, he was assisted by a member of the research team. The total time to complete the items was approximately twenty minutes, although many children finished in less time.

The Listening Measure of eighteen items was administered next using a recording of the music excerpts. Subjects marked answers on a multiple-choice answer sheet. (See Appendix C for musical source of examples included, the answer sheet used by subjects, and directions for administration.) The measure was divided into three sections, with instructions and sample items for each section. Children were encouraged to ask questions concerning the procedure after each set of instructions. Including the time for directions and sample items, the testing time was approximately twenty-five minutes.

The Manipulative Measure of eighteen items was administered individually to children who were randomly selected from the classroom samples. (See Appendix C for method of administration and a copy of test items.) The administrator asked the child to play various music-making instruments in different combinations and make judgments concerning the pitch, duration, or loudness of the sounds produced. Each of the eighteen judgments was rated right or wrong. Time for administration of the measure varied from ten to fifteen minutes.

The Overt Measure was the final measure in order of administration. It contained nine musical examples to which subjects responded in two ways: (1) by overt-physical movement, showing a change in movement when the music changed, and (2) by an oral response telling how the music changed. A general classroom orientation to the Overt Measure was included in the group-testing session, after the Verbal and Listening Measures were completed; individual subjects were again given complete instructions immediately prior to beginning the measure. (See Appendix C for classroom orientation, method of administration, copy of musical sources, judges' answer form, and stimulus card for oral answers.) The subject heard each of the items three times. He

---

[1] Beginning with the third school system tested, a chart of seven words on which children most frequently requested help was read for and by the children before the administration of the measure (Appendix C).

was instructed to (1) listen to the music the first time and decide how he could move to it, (2) move with the music on the second hearing, showing a change where the music changed, and (3) listen again, refer to the card if necessary, and tell what change occurred. The administrator checked the answer form, giving one point for change of movement at the proper time and one point for the correct oral answer. Testing time was approximately fifteen minutes per subject.

Results

Reliability

The means and standard deviations of the Verbal, Listening, Manipulative, and Overt Measures appear in Table XVII. Since the Verbal and Listening Measures were administered to the total sample of 429 subjects, means and standard deviations are presented for both the total sample of 429 and the random sample of 214 subjects. It should be noted that the means and standard deviations for the Listening Measure were computed on the eighteen items, with each item, including those with double answers, receiving one point. In other analyses dividing the Listening Measure into scores for pitch, duration, and loudness (Tables XVIII, XIX, and XX), double items were scored for two dimensions, producing eight points for each dimension for a total possible score of twenty-four.

TABLE XVII

MEANS AND STANDARD DEVIATIONS
FOR THE FOUR MEASURES
MAIN STUDY

Measure	Number of items	Number of subjects	Mean	Standard deviation
Verbal	18	214	8.39	3.54
Verbal	18	429	8.26	3.51
Listening	18	214	9.43	3.80
Listening	18	429	9.18	3.76
Manipulative	18	214	10.98	2.94
Overt	18	214	10.43	2.93

The means and standard deviations of the derived scores for the dimensions of pitch, duration, and loudness, as measured by the four separate measures, appear in Table XVIII. The means and standard deviations for the dimensions of pitch, duration, and loudness, derived from the combined Written Measures and Nonwritten Measures, appear in Table XIX.

TABLE XVIII

MEANS AND STANDARD DEVIATIONS FOR THE DERIVED SCORES FROM THE FOUR MEASURES
MAIN STUDY
N = 214

Measure	Number of items	Dimension					
		Pitch		Duration		Loudness	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
Verbal	6	2.486	1.400	3.206	1.599	2.706	1.460
Listening	8	3.818	2.122	5.313	1.772	5.[illegible]9	1.561
Manipulative	6	2.276	1.720	4.154	1.360	4.551	1.104
Overt	6	1.930	1.346	4.439	1.247	4.065	1.341

TABLE XIX

MEANS AND STANDARD DEVIATIONS OF THE WRITTEN AND NONWRITTEN DERIVED SCORES
MAIN STUDY
N = 214

Derived score	Number of items	Mean	Standard deviation
Written Pitch	14	6.31	3.01
Nonwritten Pitch	12	4.21	2.42
Written Duration	14	8.52	2.84
Nonwritten Duration	12	8.59	2.06
Written Loudness	14	8.21	2.44
Nonwritten Loudness	12	8.62	1.79

TABLE XX

STRATA FIXED RELIABILITY, STRATA RELIABILITY, RATIO OF SUBJECTS BY STRATA MEAN SQUARE TO RESIDUAL MEAN SQUARE, AND CORRELATIONS AMONG STRATA FOR THE FOUR MEASURES
MAIN STUDY

Measure	Number of Subjects	Strata Fixed Reliability	Strata Reliability			Ratio of Ss by Strata M.S. to Residual M.S.	Correlation Among Strata		
			P	D	L		r_{PxD}	r_{PxL}	r_{DxL}
Verbal (18 items)	214	.708	.389	.525	.447	1.061*	.387*	.448*	.450*
Verbal (18 items)	429	.700	.440	.476	.415	1.036*	.431*	.415*	.433*
Listening (18 items-24 points)	214	.845	.678	.639	.538	.753	.725*	.668*	.766*
Listening (18 items-24 points)	429	.846	.674	.647	.549	.776	.726*	.678*	.735
Manipulative (18 items)	214	.656	.671	.438	.210	1.579*	.338*	.223*	.070
Overt (18 items)	214	.644	.416	.453	.371	1.145*	.258*	.326*	.396

*Significant at .05 level

The reliability estimates, ratio of subject by strata mean square to residual mean square, correlations among strata, and strata reliabilities were computed for the four measures using an analysis of variance computer program, RELIB (Appendix E). These data appear in Table XX.

Table XX shows that the reliability estimates range from .644 to .846, with the Written Measures yielding the highest reliability estimates. Most of the strata reliabilities are substantial; however, the strata reliabilities for the Listening Measure are higher than those for the other three measures. The ratio of subjects by strata mean square to residual mean square (test of the hypothesis that strata do not exist) was not found to be significant for the Listening Measure. Further evidence of nonexistent strata in the Listening Measure is shown by item correlations between strata as high as item correlations within strata. The correlations among strata range from .07 to .45 for the remaining measures, and each correlation, except the correlation between Manipulative Duration and Loudness, is significantly different from zero.

Table XXI presents reliability estimates for written and nonwritten derived scores, strata reliabilities, ratio of subject by strata mean square to residual mean square, and correlations among strata, computed by a computer analysis of variance program, RELIB (Appendix E).

An examination of the ratio of subjects by strata mean square shows that strata exist for each of the derived scores. This ratio indicates that the Verbal and Listening Measures are separate measures contributing different information to the written derived scores. This is also true of the Overt and Manipulative Measures.

The correlations among strata, ranging from .322 to .426, are further evidence that the Verbal and Listening Measures contribute different information to the written derived scores. The Manipulative and Overt Measures, with correlations ranging from .061 to .246, contribute different information to the nonwritten derived scores.

Item Analysis

An item analysis was computed for each measure, using a library computer program ITANL (Appendix E). Tables XXII to XXXIII present the item analyses for the four measures (Verbal, Listening, Manipulative, and Overt) and for the written and nonwritten derived scores for the dimensions of pitch, duration, and loudness. These include the following information: per cent of subjects answering an item correctly (item difficulty index), item-test correlation (item discrimination index), and the dimension measured.

The item analyses for the Verbal and Listening Measures were computed for both the random sample of 214 subjects and for the total sample of 429; therefore two item analysis tables appear for each of these measures (Tables XXII to XXV). Since the Manipulative and Overt Measures were administered only to the random sample of 214, a single item analysis appears for each measure (Tables XXVI and XXVII).

TABLE XXI

STRATA FIXED RELIABILITY, STRATA RELIABILITY, RATIO OF SUBJECTS BY STRATA MEAN SQUARE TO RESIDUAL MEAN SQUARE, AND CORRELATIONS BETWEEN STRATA FOR THE WRITTEN AND NONWRITTEN DERIVED SCORES MAIN STUDY

Derived Scores	Number of Subjects	Strata fixed Reliability	Strata Reliabilities		Ratio of Ss by Strata M.S. to Residual M.S.	Correlation Between Strata
			Verbal	Listening		
Written Pitch	214	.706	.380	.682	1.308	.426
Written Pitch	429	.725	.432	.673	1.167	.499
Written Duration	214	.689	.525	.593	1.390	.405
Written Duration	429	.657	.474	.548	1.252	.413
Written Loudness	214	.591	.437	.477	1.297	.322
Written Loudness	429	.574	.412	.448	1.220	.333
			Manipulative	Overt		
Nonwritten Pitch	214	.652	.675	.414	1.838	.226
Nonwritten Duration	214	.553	.438	.451	1.357	.246
Nonwritten Loudness	214	.345	.210	.371	1.354	.061

TABLE XXII

VERBAL MEASURE ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Dimension measured
1	.720	.379	L
2	.350	.371	P
3	.360	.484	D
4	.617	.587	D
5	.589	.544	D
6	.322	.629	L
7	.318	.498	L
8	.439	.522	D
9	.164	.381	P
10	.593	.571	D
11	.519	.518	L
12	.463	.648	L
13	.407	.373	P
14	.364	.481	L
15	.579	.561	P
16	.612	.583	D
17	.570	.527	P
18	.421	.666	P

TABLE XXIII

VERBAL MEASURE ITEM ANALYSIS
MAIN STUDY
N = 429

Item	Per cent of students answering correctly	Item-test correlation	Dimension measured
1	.718	.427	L
2	.366	.390	P
3	.406	.474	D
4	.599	.554	D
5	.543	.589	D
6	.336	.503	L
7	.275	.398	L
8	.466	.512	D
9	.177	.488	P
10	.534	.505	D
11	.541	.541	L
12	.417	.641	L
13	.417	.408	P
14	.368	.486	L
15	.541	.569	P
16	.606	.552	D
17	.564	.568	P
18	.389	.649	P

TABLE XXIV

LISTENING MEASURE ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Dimension measured
1	.500	.599	L
2	.533	.758	P
3	.542	.739	P
4	.369	.590	L
5	.360	.742	D
6	.855	.483	D
7	.556	.641	D + L
8	.607	.582	P + L
9	.626	.430	D + L
10	.393	.863	P + D
11	.621	.678	P + D
12	.313	.613	P + L
13	.897	.571	D
14	.463	.365	P
15	.575	.450	L
16	.164	.295	P
17	.626	.423	L
18	.458	.518	D

TABLE XXV

LISTENING MEASURE ITEM ANALYSIS
MAIN STUDY
N = 429

Item	Per cent of students answering correctly	Item-test correlation	Dimension measured
1	.487	.665	L
2	.462	.755	P
3	.473	.694	P
4	.357	.584	L
5	.368	.749	D
6	.862	.441	D
7	.557	.675	D + L
8	.594	.639	P + L
9	.615	.427	D + L
10	.368	.872	P + D
11	.590	.678	P + D
12	.331	.605	P + L
13	.893	.465	D
14	.410	.340	P
15	.594	.392	L
16	.156	.301	P
17	.646	.366	L
18	.431	.554	D

TABLE XXVI

MANIPULATIVE MEASURE ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Dimension measured
1	.477	.464	D
2	.612	.522	D
3	.308	.671	P
4	.280	.600	P
5	.780	.399	L
6	.748	.396	L
7	.570	.626	P
8	.579	.722	P
9	.612	.555	D
10	.813	.419	D
11	.813	.333	L
12	.794	.310	L
13	.285	.665	P
14	.252	.654	P
15	.757	.399	D
16	.879	.398	D
17	.505	.249	L
18	.911	.347	L

TABLE XXVII

OVERT MEASURE ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Dimension measured
1	.832	.450	L
2	.860	.691	D
3	.893	.576	D
4	.505	.576	L
5	.855	.628	D
6	.178	.441	P
7	.145	.316	P
8	.706	.572	L
9	.542	.710	P
10	.617	.516	L
11	.850	.365	D
12	.491	.570	D
13	.766	.444	L
14	.491	.398	D
15	.369	.340	P
16	.449	.517	P
17	.640	.428	L
18	.248	.493	P

Note: Items 1-9 are Overt Responses
Items 10-18 are Oral Responses to the same musical examples

TABLE XXVIII

WRITTEN PITCH ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Original item number
1	.350	.379	2
2	.164	.484	9
3	.411	.386	13
4	.579	.657	15
5	.565	.491	17
6	.421	.528	18
7	.533	.861	2
8	.542	.771	3
9	.659	.533	8
10	.467	.855	10
11	.640	.640	11
12	.346	.713	12
13	.463	.408	14
14	.168	.261	16

Note: Items 1-6 are from Verbal Measure
Items 7-14 are from Listening Measure

TABLE XXIX

WRITTEN DURATION ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Original item number
1	.360	.507	3
2	.617	.661	4
3	.589	.532	5
4	.439	.582	8
5	.593	.599	10
6	.612	.529	16
7	.360	.669	5
8	.855	.433	6
9	.640	.540	7
10	.701	.444	9
11	.598	.670	10
12	.790	.616	11
13	.897	.563	13
14	.458	.620	18

Note: Items 1-6 are from Verbal Measure
Items 7-14 are from Listening Measure

TABLE XXX

WRITTEN LOUDNESS ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Original item number
1	.720	.428	1
2	.322	.532	6
3	.322	.612	7
4	.519	.601	11
5	.458	.482	12
6	.364	.459	14
7	.500	.725	1
8	.369	.542	4
9	.888	.584	7
10	.925	.402	8
11	.734	.567	9
12	.874	.162	12
13	.575	.496	15
14	.621	.500	17

Note: Items 1-6 are from Verbal Measure
Items 7-14 are from Listening Measure

TABLE XXXI

NONWRITTEN PITCH ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Original item number
1	.308	.606	3
2	.280	.577	4
3	.575	.686	7
4	.579	.721	8
5	.285	.650	13
6	.252	.792	14
7	.178	.380	6
8	.369	.334	7
9	.145	.390	9
10	.449	.624	15
11	.537	.461	16
12	.252	.661	18

Note: Items 1-6 are from Manipulative Measure
Items 7-12 are from Overt Measure

TABLE XXXII

NONWRITTEN DURATION ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Original item number
1	.477	.615	1
2	.612	.596	2
3	.612	.514	9
4	.813	.463	10
5	.757	.590	15
6	.879	.494	16
7	.860	.565	2
8	.850	.583	3
9	.888	.592	5
10	.491	.553	11
11	.855	.598	12
12	.495	.487	14

Note: Items 1-6 are from Manipulative Measure
Items 7-12 are from Overt Measure

TABLE XXXIII

NONWRITTEN LOUDNESS ITEM ANALYSIS
MAIN STUDY
N = 214

Item	Per cent of students answering correctly	Item-test correlation	Original item number
1	.780	.518	5
2	.748	.438	6
3	.813	.420	11
4	.794	.413	12
5	.505	.375	17
6	.911	.255	18
7	.832	.391	1
8	.617	.577	4
9	.505	.486	8
10	.766	.559	10
11	.706	.554	13
12	.640	.455	17

Note: Items 1-6 are from Manipulative Measure
Items 7-12 are from Overt Measure

Item analyses for the Verbal and Listening Measures yielded better discrimination and difficulty indices than did analyses for the Manipulative and Overt Measures. In examining the item analyses for the written and nonwritten measurements of the separate dimensions (Tables XXVIII to XXXIII), it is evident that Written Pitch, Written Duration, and Written Loudness, according to the obtained indices, contain more effective items than Nonwritten Pitch, Nonwritten Duration, and Nonwritten Loudness.

Correlations and Partial Correlations

Table XXXIV presents a correlation matrix of the scores from the four measures, a standard intelligence test, and a standard reading achievement test. The intercorrelations of the four measures are approximately the same as correlations of the four measures with IQ and with reading achievement. Among the intercorrelations of the four measures, the Listening Measure correlates the highest with each of the other three measures. Verbal and Listening Measures correlate higher with IQ and reading scores than do Manipulative and Overt.

Table XXXV presents an intercorrelation matrix of the written and nonwritten derived scores and IQ and reading scores. Various sized samples are presented in the matrix, since the school districts involved had used different IQ and reading tests; as is evident from the table, the correlations did not change appreciably from sample to sample. Each of the written derived scores for pitch, duration, and loudness correlates higher with reading scores than written derived scores correlate with nonwritten derived scores. The written derived scores for each dimension correlate higher with IQ and reading scores than the nonwritten derived scores correlate with IQ and reading scores.

Table XXXVI presents correlations between the written and nonwritten derived scores, partial correlations holding out IQ, and partial correlations holding out reading scores. These partial correlations yield an estimate of the amount of correlation with the IQ or reading variable held constant. The difference between the partial and the original correlations suggests the amount of correlation due to the influence of reading or IQ. Significance of the partial correlations was computed using a t test (McNemar, 34).

The partial correlations between Written and Nonwritten Pitch and Written and Nonwritten Duration were significant after either IQ or reading scores were partialled out, indicating that there was variance in these scores attributable to the concepts of pitch and duration. The partial correlation between Written and Nonwritten Loudness was not significant, indicating that the original correlation between the two measurements may have existed because of their relationship to either IQ or reading. This lack of relationship may be due in part to the low reliability of Nonwritten Loudness.

The means and standard deviations for the IQ and reading scores appear in Table XXXVII. As indicated, an unrestricted range for the IQ and reading scores entered into the correlations.

TABLE XXXIV

SUMMARY CORRELATION MATRIX FOR THE FOUR MEASURES, IQ
AND READING FOR VARIOUS SELECTED SAMPLES
MAIN STUDY

	Verbal				Listening				Manipulative				Overt			
Measure	N=57	N=94	N=69	N=214	N=57	N=94	N=69	N=214	N=57	N=94	N=69	N=214	N=57	N=94	N=69	N=214
Verbal																
Listening	.671	.590	.676	.593												
Manipulative	.468	.372	.479	.383	.513	.458	.523	.460								
Overt	.509	.498	.538	.455	.632	.586	.652	.526	.378	.264	.427	.318				
IQ	.621	.598			.596	.574			.456	.462			.425	.429		
Reading	.692		.666		.624		.640		.520		.504		.344		.362	

N = 57 California Mental Maturity and California Achievement
N = 94 California Mental Maturity
N = 69 California Achievement
N = 214 Appropriate data not available in terms of IQ or Reading Scores.

TABLE XXXV

SUMMARY CORRELATION MATRIX FOR THE WRITTEN AND NONWRITTEN DERIVED SCORES, IQ, AND READING FOR VARIOUS SELECTED SAMPLES*
MAIN STUDY

	Written Pitch				Written Duration				Written Loudness				Nonwritten Pitch				Nonwritten Duration				Nonwritten Loudness			
	N= 214	N= 94	N= 69	N= 57	N= 214	N= 94	N= 69	N= 57	N= 214	N= 94	N= 69	N= 57	N= 214	N= 94	N= 69	N= 57	N= 214	N= 94	N= 69	N= 57	N= 214	N= 94	N= 69	N= 57
W.P.	.71																							
W.D.	.62	.65	.74	.75	.69																			
W.L.	.55	.52	.61	.56	.59	.61	.65	.64	.59															
N.P.	.48	.55	.59	.59	.41	.53	.60	.63	.39	.42	.53	.50	.65											
N.D.	.42	.46	.38	.35	.43	.47	.52	.48	.39	.41	.45	.42	.40	.43	.40	.36	.55							
N.L.	.39	.40	.49	.51	.31	.39	.43	.49	.34	.33	.43	.34	.35	.39	.58	.57	.33	.31	.38	.41	.35			
IQ		.50		.53		.67		.74		.52		.52		.44		.46		.48		.42		.35		.37
READ			.66	.66			.66	.67			.60	.59			.50	.52			.33	.32			.36	.34

* IQ measured by California Mental Maturity
Reading measured by California Achievement

Note: The self-correlation is the reliability estimate from the RELIB computer program.

TABLE XXXVI

WRITTEN AND NONWRITTEN DERIVED SCORES, CORRELATIONS AND PARTIAL CORRELATIONS HOLDING OUT IQ AND READING*
MAIN STUDY
N = 57

	Written Pitch			Written Duration			Written Loudness			Nonwritten Pitch			Nonwritten Duration		
	r	Partial r holding out: IQ	Read	r	Partial r holding out: IQ	Read	r	Partial r holding out: IQ	Read	r	Partial r holding out: IQ	Read	r	Partial r holding out: IQ	Read
W.D.	.747	.621	.564												
W.L.	.562	.335	.305	.641	.451	.414									
N.P.	.590	.354	.504	.631	.490	.444	.498	.341	.274						
N.D.	.354	.170	.210	.478	.277	.378	.421	.262	.307	.355	.202	.235			
N.L.	.509	.394	.392	.491	.342	.166	.343	.184	.176	.574	.489	.387	.409	.298	.281

*Values below .263 are nonsignificant at .05 level.

TABLE XXXVII

MEANS AND STANDARD DEVIATIONS FOR READING AND IQ SCORES, SELECTED SAMPLES MAIN STUDY

Standard test	Number of subjects	Mean	Standard deviation
IQ	57	107.19	13.87
IQ	94	105.79	13.16
Reading	57	3.89	1.01
Reading	69	4.00	.95

Validity

Content validity was investigated and established in the development of the Battery of Musical Concept Measures by consensus of experts. This is explained in the Development of the Measures (Chapter III), and is discussed further in the discussion section of this chapter.

Construct validity was investigated in the Main Study in relation to rigorous criteria set forth by Campbell and Fiske (9), who propose evaluating a multitrait-multimethod matrix by four criteria:

1. "The entries in the validity diagonal should be significantly different from zero and sufficiently large to encourage further examination of validity. This requirement is evidence of convergent validity" (p. 81). In Table XXXVIII the validity estimates for pitch, duration, and loudness are respectively .481, .427, and .343, all significantly different from zero.

2. "A validity diagonal value should be higher than the value lying in its column and row in the heterotrait-hetermethod triangles" (p. 81). In Table XXXVIII the validity estimates for pitch and duration, but not loudness, are higher than the correlations in the heterotrait-heteromethod triangle.

3. "A variable /should/ correlate higher with an independent effort to measure the same trait than with measures designed to get at different traits which happen to employ the same method" (p. 82). In Table XXXVIII evidence of discriminant validity is given for Nonwritten Duration and Nonwritten Pitch. The highest correlations obtained are between Written Measures, regardless of the concept being measured, showing the presence of method variance in the correlations.

4. "The same pattern of trait interrelationship /should/ be shown in all of the heterotrait triangles of both the monomethod and heteromethod blocks" (p. 82). In Table XXXVIII this desideratum is met and provides evidence for discriminant validity.

TABLE XXXVIII

CORRELATION MATRIX FOR THE WRITTEN AND NONWRITTEN MEASURES
(MULTITRAIT-MULTIMETHOD MATRIX)
MAIN STUDY
N = 214

Derived scores	W.P.	W.D.	W.L.	N.P.	N.D.	N.L.
Written Pitch	.706					
Written Duration	.620	.689				
Written Loudness	.547	.587	.591			
Nonwritten Pitch	.481	.412	.394	.652		
Nonwritten Duration	.417	.427	.386	.398	.553	
Nonwritten Loudness	.394	.310	.343	.348	.326	.345

Secondary Analysis of the Overt Measure

Table XXXIX presents the percentages of oral responses for the Overt Measure. For each item, any one of the six responses (higher, lower, faster, slower, louder, softer) was possible since the items were open-ended with forced choice of one of the given responses. The obtained responses were tallied to investigate possible confusion of labels. For example, it had been expected that lower might be given as a response to softer music, and higher as a response to louder music, more often than other incorrect responses.

Table XXXIX shows that lower was the most frequent incorrect response to softer music, and higher was the most frequent incorrect response to louder music. When the music became faster, higher and louder were the incorrect responses equally chosen. When the music became slower, softer was the most frequent incorrect response, chosen by at least 26 per cent of the subjects. For items in which the musical change was higher, no consistent trend was found in the incorrect responses, except that lower and slower were the responses least chosen. Since the items measuring a musical change of higher resulted in a wide variety of responses, implications from these items and the one lower item have not been drawn.

The Overt Measure contained nine items, each item having an overt response and an oral response. An estimate of the agreement between the oral and overt responses for each item was computed by Kuder-Richardson Formula 21, using as an agreement score for each subject the total number of times the subject had the same response, correct or incorrect, on both parts of an item (e.g., a score of 0 indicated no agreement on responses and a score of 3 indicated agreement on all response pairs for each dimension). As indicated in Table XL, the agreement between the overt and oral responses for each of the dimensions of pitch, duration, and loudness was low and nonsignificant. The means for the pitch, duration, and loudness agreement scores were approximately the same; this was also true for the standard deviations. There was no high

agreement between the overt and oral responses on any dimension. Children's movements to elements of music, as investigated in this study, did not correlate with what they expressed orally as having happened in the music.

TABLE XXXIX

ORAL RESPONSES GIVEN TO
OVERT MEASURE ITEMS
MAIN STUDY
N = 214

Item number	Correct answer	Response					
		Higher	Lower	Louder	Softer	Faster	Slower
1	Softer	.08	.15	.03	.62	.01	.10
4	Louder	.13	.00	.77	.01	.10	.00
8	Louder	.14	.04	.64	.02	.12	.02
2	Faster	.06	.004	.06	.02	.85	.004
3	Slower	.00	.11	.01	.37	.01	.50
5	Slower	.06	.12	.03	.26	.03	.49
6	Higher	.37	.08	.24	.07	.16	.07
9	Higher	.25	.08	.07	.21	.28	.09
7	Lower	.06	.45	.08	.11	.20	.09

TABLE XL

MEANS, STANDARD DEVIATIONS, AND KUDER-RICHARDSON
ESTIMATES OF AGREEMENT FOR THE OVERT AND
ORAL RESPONSES ON THE OVERT MEASURE
MAIN STUDY
N = 214

Dimension	Mean	Standard deviation	Estimate of agreement*
Pitch	1.640	.868	.023
Duration	1.804	.885	.123
Loudness	1.654	.923	.192

*Kuder-Richardson Formula No. 21

Stanines

Nonnormalized stanine values were obtained for Written and Nonwritten Pitch, Duration, and Loudness (Tables XLI to L). A stanine value was obtained from each possible score for each dimension on each measure. Since stanines are whole numbers by definition, each stanine value was rounded to the nearest whole number by the following procedure:

> If a value was less than or equal to X.50, this value was made equal to X. If an obtained value was greater than X.50, this value was made equal to X + 1. If a computed stanine was less than 1 or greater than 9, this value was made equal to 1 or 9, since stanines are, by definition, not less than 1 or greater than 9.

The Verbal and Listening stanines were averaged to obtain a written score for each dimension. The Manipulative and Overt stanines were averaged to obtain a nonwritten score for each dimension. In averaging these stanines, any value equal to or greater than X.50 was made equal to X + 1 (e.g., $\frac{2 + 3}{2}$ = 2.5 = 3).

Discussion

Reliability of Measures (Tables XX and XXI, pp. 50 and 52)

One of the most important characteristics of a test is reliability. "Reliability always refers to consistency throughout a series of measurements" (Cronbach, 11, p. 126). Reliability estimates can be obtained using statistics applied to the techniques of test-retest (stability estimate), parallel forms of the same test (equivalency estimate), or single administration of the test (internal consistency estimate). Each type of reliability yields a different type of information, but each yields an estimate of the amount of test variance that is not error variance.

Reliability for the Battery of Musical Concepts was investigated using an internal consistency estimate. A strata-fixed reliability estimate was chosen for the measurement of internal consistency, because the measures were heterogeneous and the strata fixed rather than random.

The Listening Measure presents the highest overall reliability (.85) as well as the highest reliabilities for the three dimensions; in terms of reliability, it is the most satisfactory of the four measures. The reliability of the Verbal Measure (.71) is satisfactory for group measurement (Thorndike and Hagen, 52, p. 190).

The overall reliability estimate for the Manipulative Measure (.66) is comparable to the estimate for the Overt Measure (.64). However, the strata reliabilities in the Manipulative Measure are more variable than the strata reliabilities in the Overt Measure. The low reliability estimate for Manipulative Loudness (.21), the lowest of the strata reliabilities, may be due in part to (1) the fine discriminations required, (2) the individual

TABLE XLI

STANINE CONVERSION TABLE
WRITTEN PITCH

Number correct on Listening Pitch	Number correct on Verbal Pitch						
	0	1	2	3	4	5	6
0	1	2	3	4	4	5	6
1	2	3	3	4	5	6	6
2	2	3	4	5	5	6	7
3	3	4	4	5	6	7	7
4	3	4	5	6	6	7	8
5	4	5	5	6	7	8	8
6	4	5	6	7	7	8	9
7	5	6	6	7	8	9	9
8	5	6	7	8	8	9	9

TABLE XLII

STANINE CONVERSION TABLE
NONWRITTEN PITCH

Number correct on Overt Pitch	Number correct on Manipulative Pitch						
	0	1	2	3	4	5	6
0	2	3	4	4	5	5	6
1	3	4	5	5	6	6	7
2	4	5	5	6	6	7	7
3	5	6	6	7	7	8	8
4	5	6	7	7	8	8	9
5	6	7	7	8	8	9	9
6	6	7	7	8	8	9	9

TABLE XLIII

STANINE CONVERSION TABLE
WRITTEN DURATION

Number correct on Listening Duration	Number correct on Verbal Duration						
	0	1	2	3	4	5	6
0	1	2	2	3	4	4	5
1	1	2	2	3	4	4	5
2	1	2	2	3	4	4	5
3	2	2	3	4	4	5	5
4	3	3	4	5	5	6	6
5	3	4	4	5	6	6	7
6	4	4	5	6	6	7	7
7	4	5	5	6	7	7	8
8	5	5	6	7	7	8	8

TABLE XLIV

STANINE CONVERSION TABLE
NONWRITTEN DURATION

Number correct on Overt Duration	Number correct on Manipulative Duration						
	0	1	2	3	4	5	6
0	1	1	2	2	3	4	5
1	1	1	2	2	3	4	5
2	1	1	2	2	3	4	5
3	2	2	3	3	4	5	6
4	3	3	3	4	5	5	6
5	4	4	4	5	6	6	7
6	4	4	5	5	6	7	8

TABLE XLV

STANINE CONVERSION TABLE
WRITTEN LOUDNESS

Number correct on Listening Loudness	Number correct on Verbal Loudness						
	0	1	2	3	4	5	6
0	1	2	3	3	4	5	5
1	1	2	3	3	4	5	5
2	1	2	3	3	4	5	5
3	2	3	3	4	5	5	6
4	2	3	4	4	5	6	6
5	3	4	4	5	6	6	7
6	4	5	5	6	7	7	8
7	4	5	6	6	7	8	8
8	5	6	6	7	8	8	9

TABLE XLVI

STANINE CONVERSION TABLE
NONWRITTEN LOUDNESS

Number correct on Overt Loudness	Number correct on Manipulative Loudness						
	0	1	2	3	4	5	6
0	1	1	1	2	3	4	5
1	1	1	1	2	3	4	5
2	2	2	2	2	3	4	5
3	2	2	2	3	4	5	6
4	3	3	3	4	5	6	7
5	4	4	4	4	5	6	7
6	5	5	5	5	6	7	8

TABLE XLVII

STANINE CONVERSION TABLE
VERBAL MEASURE

Number correct	Dimension		
	Pitch	Duration	Loudness
0	1	1	1
1	3	2	3
2	4	3	4
3	6	5	5
4	7	6	7
5	9	7	8
6	9	8	9

TABLE XLVIII

STANINE CONVERSION TABLE
LISTENING MEASURE

Number correct	Dimension		
	Pitch	Duration	Loudness
0	1	1	1
1	2	1	1
2	3	1	1
3	4	2	2
4	5	4	3
5	6	5	4
6	7	6	6
7	8	7	7
8	9	8	8

TABLE XLIX

STANINE CONVERSION TABLE
MANIPULATIVE MEASURE

Number correct	Dimension		
	Pitch	Duration	Loudness
0	2	1	1
1	4	1	1
2	5	2	1
3	6	3	2
4	7	5	4
5	8	6	6
6	9	8	8

TABLE L

STANINE CONVERSION TABLE
OVERT MEASURE

Number correct	Dimension		
	Pitch	Duration	Loudness
0	2	1	1
1	4	1	1
2	5	1	2
3	7	3	3
4	8	4	5
5	9	6	6
6	9	7	8

differences in manipulation of the devices, which resulted in a variety of acceptable answers, and (3) the large proportion of subjects making correct responses to the items, or a combination of these.

In general, the Verbal and Listening Measures are superior to the Overt and Manipulative Measures, both in total reliability and in reliability for the three dimensions.

The ratio of subjects by strata mean square to residual mean square is a test of the hypothesis that strata do not exist. All ratios for the three dimensions on the Written Measures are significant, indicating that the Verbal and Listening Measures are separate measures contributing different information to the written derived scores. The ratios for the Nonwritten Measures indicate that the Manipulative and Overt Measures contributed different information to the nonwritten derived scores. This provides evidence that the Verbal, Listening, Manipulative, and Overt Measures are separate measures, measuring either different aspects of the musical concepts or different concepts. The authors believe that these four modes of measurement reveal different aspects of the musical concepts of pitch, duration, and loudness.

Correlations among strata are further indication of the relationships among the different modes of measurement. Generally, a high correlation may indicate that the strata are contributing similar information; a low correlation may indicate that the measures are different and are contributing unique information in that they are measuring different aspects of the same concepts, or that they contributed different information by measuring different concepts. Low reliability in separate measures also may contribute to low correlations.

The correlations among strata, ranging from .322 to .426 for the Written Measures, suggest that the Verbal and Listening Measures are contributing different information to the written derived scores for the different dimensions. The correlations among strata for the Nonwritten Measures, ranging from .061 to .246 for the different dimensions, suggest that the Manipulative and Overt Measures contributed different information to the nonwritten derived scores. The authors believe that the low correlations among strata are evidence that the Verbal Measure is different from the Listening Measure in measuring concepts of pitch, duration, and loudness, and the Manipulative Measure is different from the Overt Measure in measuring these concepts.

Item Analysis (Tables XXII to XXXIII, pp. 53 to 64)

Item analyses are used primarily for selecting the best items for the final form of a test, also for improving validity and reliability. In general, items with item-test correlations between .30 and .80 present tests with satisfactory reliability and validity estimates, and difficulty indices of approximately .50 produce favorable reliability estimates (Guilford, 18, p. 471).

Of the thirty-six items in the Verbal and Listening Measures (Tables XXII to XXV, pp. 53 to 56), thirty are acceptable according to the stated criteria, contributing to the favorable reliability estimates of these

measures. It is assumed that future revision of the remaining six items, to conform with the criteria, would further increase the reliability.

In the Manipulative Measure (Table XXVI, p. 57) items measuring the pitch dimension are more acceptable, according to the criteria, than items measuring duration or loudness. One reason for this is that most subjects made correct responses to the duration and loudness items.

The Overt Measure item analysis (Table XXVII, p. 58) shows better indices for the oral responses than for the overt responses, with only three of the nine overt responses meeting the difficulty criterion. This is indication that the oral responses constituted better items than the overt responses.

In summary, the written items for pitch, duration, and loudness (derived from the Verbal and Listening Measures) are generally superior to the nonwritten items (derived from the Manipulative and Overt Measures). This is also apparent in the item analyses for the derived scores for each dimension (Tables XXVIII to XXXIII, p. 59 to 64). The implication is that the Written Measures provided the more effective modes for identifying children's musical concepts.

Correlations and Partial Correlations (Tables XXXIV to XXXVII, pp. 66 to 69)

The present study investigated the correlation of the Battery of Musical Concept Measures with a measure of intelligence, in accordance with the following statement of Campbell (8):

> A new test, no matter what its content, should be correlated with an intelligence test of as similar format as possible. . . . If correlations are reported with independent trait-appropriate or criterion measures, it should be demonstrated that the new test correlates better with these measures than does the intelligence test (p. 548).

Correlations were computed for the written and nonwritten derived scores of each dimension with IQ scores (California Test of Mental Maturity) and reading achievement scores (California Achievement Test) for all subjects with these scores available. Because of the relatively small samples (fifty-seven for both scores, ninety-four for IQ, and sixty-nine for reading), conclusions drawn from this data must be considered tentative (Table XXXV, p. 67).

The correlations for Written Pitch, and for Nonwritten Duration in one instance, meet Campbell's recommendation; the correlations for the other derived scores do not. This may be due in part to the higher reliability of the standardized IQ test entering into the correlations with the derived scores as opposed to the lower reliabilities of the two derived scores being correlated. The implication is that the measures should be further refined, with efforts concentrated on increasing their reliability and validity. All correlations of both IQ and reading scores with written derived scores are higher than correlations of IQ and reading scores with nonwritten derived scores. A factor of general mental ability seems to be reflected in all

correlations, since none is zero or negative. This tends to confirm the authors' opinion that measures of musical concepts which incorporate cognitive aspects of musical understanding almost inevitably will show a substantial correlation with intelligence.

The correlation of reading scores with Overt Measure scores, while significant, is lower than reading score correlations with the other three measures, accounting for only a small amount of the variance. This suggests that performance on the Overt Measure is less dependent on reading ability than performance on the other measures; the Overt Measure, therefore, may be suitable for use with deficient readers.

Table XXXVI (p. 68) shows partial correlations for the written and nonwritten derived scores of the three dimensions holding out reading and IQ scores. The consistent decrease in the intercorrelations when either IQ or reading scores are partialled out suggests that IQ and reading achievement enter similarly into all of the intercorrelations, and the relationship is not unique to any particular correlation. Correlations between Written and Nonwritten Pitch, and between Written and Nonwritten Duration, are significant after partialling out either reading or IQ scores; the partial correlation between Written and Nonwritten Loudness did not reach significance. This suggests that a factor in addition to intelligence or reading achievement contributed to the variance in the pitch and duration scores. The authors believe this factor is the musical concept being measured and that concepts of pitch and duration are indeed accounting for some variance in these scores.

Validity

Validity, as Guilford (18) notes, is a "highly relative concept," its chief requisite being that, when exhibited as the quality of a test, the test measure what it is supposed to measure. Validity, however, varies with the use to which the instrument possessing it is put; in Guilford's words, the question should not be "Is this test valid?" but rather "Is it valid for what?" (p. 461)

The experimental Battery of Musical Concept Measures described in the present study began with the idea of developing a technique for identifying children's concepts of pitch, duration, and loudness, or, to paraphrase Cronbach (11), with only a concept for which a testing instrument was desired. The interpretation of tests, Cronbach observes:

> . . . is built up very gradually, and probably is never complete. As knowledge develops, we arrive at a more complete listing of the influences that affect the test score, and may be able to estimate the strength and character of each influence. At present, the interpretation of even the best-established psychological tests falls far short of the ideal (pp. 120-1).

Establishing the validity of this study's measures, therefore, has presented challenging problems, since the Battery represents an attempt to measure a cognitive aspect of an aural (musical) experience.

In general, four types of validity are discussed with reference to psychological testing: predictive, concurrent, construct, and content. Predictive and concurrent validity are criterion oriented. Predictive validity is relevant to the establishment of a predictive function in terms of a criterion. Since the present investigators were chiefly interested in developing a technique for identifying musical concepts, predictive validity was not investigated. Concurrent validity is usually established when the intent is to substitute one test, measure, or other type of judgment, for another. Since no a priori measure or criterion was available for the present Battery, concurrent validity was not attempted.

Content validity was investigated during the development of the Measures by means of expert judgment of the items developed, the content of the Measures. Such judgment was first utilized as the research team developed, considered, accepted, rejected, and tested possible items. This practice of the team continued throughout the development of the items. Content validity was also investigated in terms of consensus of other music experts. Therefore, the items developed for the Battery of Musical Concept Measures were deemed valid for the measurement of pitch, duration, and loudness (cf. Chapter III).

Since it is generally accepted that construct validity should be investigated when no criterion in the field is available (Cronbach and Meehl, 12), this was investigated using three different approaches: (1) correlations with IQ measures, (2) correlations with reading measures, and (3) the multitrait-multimethod matrix (Campbell and Fiske, 9). It should be noted that the rigorous criteria proposed by these authors may not be entirely applicable to measurement of an essentially aural concept, and has not heretofore been applied to such a measurement. Nevertheless, it was thought important to apply the criteria essentially as a means of identifying areas in which the measures might be strengthened or refined, rather than as an assessment of present status. The discussion of construct validity, therefore, in reference to the multitrait-multimethod matrix (Table XXXVIII, p. 70) will center on the four criteria of Campbell and Fiske (cf. p. 69). This matrix presents results in line with their "typical case," showing more method variance than trait variance. Campbell and Fiske comment that such results may seem more disappointing than would validity data presented as isolated values plucked from a validity diagonal, but they also state that these isolated values present a deceptive picture of validity. The multitrait-multimethod matrix presents data from which the experimenter may draw implications for further study and research, and for developing better tests rather than simply evaluating current tests.

All validity estimates in the matrix for Written and Nonwritten Measures are significant, although certain of these could be increased through test refinement. Various interpretations of the validity estimates are possible, but the authors tend to believe that the Written Measures, with their higher reliability estimates, functioned better than the Nonwritten Measures in identifying children's musical concepts.

Secondary Analysis of Overt Measure (Tables XXXIX and XL, p. 71)

One of the expectations of the study was that children might evidence confusion in identifying the various musical elements. The Overt Measure, with a free choice of alternatives for the oral response, provided data relevant to this expectation. Although the number of items was limited, some tentative conclusions may be drawn from an analysis of these data.

1. When children were presented with a musical change softer, the most frequent incorrect response was lower. This appears to be a genuine confusion in labels, possibly resulting from many experiences with the use of lower, and related words such as down, as synonymous with softer (e.g., "Turn it down.").

2. When the musical change was louder, the most frequent incorrect response was higher. It is speculated that this is a result of hearing increase in volume referred to as up or higher (e.g., "Turn it up.").

3. When the music became slower, the most frequent incorrect response was softer; the second most frequent incorrect response was lower.

4. When the music became faster, the most frequent incorrect responses were higher and louder.

The trends shown in these observations are consistent with expectations that subjects would confuse high-loud-fast and low-soft-slow.

An examination of the relationship between the physical movement response and the oral response to individual items in the Overt Measure showed low, positive, but nonsignificant agreement between the physical and oral responses for all dimensions (Table XL, p. 71). These findings, although based on a small number of items, indicate that children's movements in response to predominant changes in the music did not agree with their spoken judgments on what happened in the music.

CHAPTER V

SUMMARY, CONCLUSIONS AND IMPLICATIONS, AND RECOMMENDATIONS FOR FUTURE RESEARCH

Summary

The Problem and Its Background

The purpose of the study was to develop a technique for identifying elementary school children's concepts of the elements of pitch, duration, and loudness. (Although timbre was accepted as a musical element, the study was limited to pitch, duration and loudness.) Four measures were developed to identify and measure concepts of the elements or dimensions named, through a multimodal technique, specifically, the Verbal, Listening, Manipulative, and Overt Measures. Past tests have involved areas of music aptitude, music achievement, and music appreciation, but little has been done in the area of musical concepts, essentially cognitive in nature.

The significance of the present problem evolved from a background of publications and speeches devoting considerable attention to the structurally-based, concept-centered music curriculum. In part, such attention reflects a general search for a more effective method of communicating knowledge; it also reflects an interest in the reexamination of music as a subject with a unique disciplinary structure.

Although literature relevant to the music education of children repeatedly cites the importance of musical concepts, the music curriculum in the elementary school in practice is still based largely on subjective experience and judgment rather than empirical evidence. An apparent gap exists between the opinions of leading music educators regarding concept-centered curriculum content and the identification of a body of knowledge on children's concepts of musical elements (dimensions). A review of related research revealed few studies pertinent to the present investigation. Yet it is meaningless to advocate the concept-centered music curriculum when there is little understanding of children's concepts of the basic musical elements. Therefore, the present study has attempted to develop a means of identifying such concepts.

The multimodal technique developed in this study, referred to as the Battery of Musical Concept Measures, included two written group measures (Verbal Measure and Listening Measure) and two nonwritten individual measures (Manipulative Measure and Overt Measure). Each of the measures employed a different stimulus-response combination to measure children's concepts of pitch, duration, and loudness.

Description of Battery of Musical Concept Measures

The Verbal Measure (reliability = .71) consisted of eighteen multiple-choice items, six items measuring each of the dimensions. This measure involved comparison and discrimination of natural or music-related sounds recalled from experience. A minimal ability to read and comprehend, as well as a background of experience with natural and musical sounds, was required for the items in this measure.

The final form of the Verbal Measure evolved from several test-and-revise cycles between July of 1965 and August of 1966. Approximately one hundred verbal items were devised, proposed, and evaluated by members of the research team. After considerable discussion of content and form, the five-answer multiple-choice format was adopted.

Although the appropriateness of a verbal mode for measuring musical concepts, aural in nature, is debatable, the consensus of the authors is that such a mode can effectively reveal musical concepts. The measure has proved effective for group measurement.

The Listening Measure (reliability = .85) was developed to measure the subject's ability to identify predominant changes in the dimensions of pitch, duration, and loudness within the multidimensional frame of reference of orchestral music. The final form of this measure consisted of eighteen short musical excerpts, four to twenty seconds in length, from standard orchestral literature. These items were of three types:

1. In the first group of six items, each contained a predominant change in one dimension, requiring a single judgment of the predominant change within the musical example, e.g., faster, lower, softer.

2. In the second group of six items, each contained predominant changes in two dimensions, requiring a judgment of two changes within the example, e.g., higher and louder, faster and softer, lower and slower.

3. In the third group of six items, each pair of musical excerpts differed predominantly in one dimension, requiring judgment of the predominant manner in which the second excerpt differed from the first.

In the development of the Listening Measure, a large number of orchestral excerpts were identified as examples having definite change in a single dimension or in two dimensions. These excerpts were analyzed and discussed as possible items. After numerous evaluation sessions, administration trials, and revisions on the basis of item analyses, the final form of the eighteen-item measure evolved.

The Manipulative Measure (reliability = .66) was developed for use with individual pupils and provided a mode for demonstrating understanding of pitch, duration, and loudness concepts through manipulation of simple music-making devices (e.g., triangle, resonator bells, finger cymbal). Items were devised to measure these concepts, and promising items were approved by consensus of the research team.

After several field trials and revisions, the final form of the eighteen-item Manipulative Measure evolved with six items measuring each of the dimensions of pitch, duration, and loudness. Items were so designed that the manual dexterity required for performing each task was minimal.

The Overt Measure (reliability = .64) was developed to provide individual subjects with a mode for demonstrating understanding of pitch, duration, and loudness concepts by responding with bodily movements and oral responses to excerpts from orchestral literature. Musical examples thirteen to twenty-five seconds in length, each containing a predominant change in one of the musical dimensions being measured, were identified as items.

The final form of the Overt Measure consisted of nine musical excerpts, with one point each scored for overt and oral responses, a total of eighteen points. Each excerpt was played three times; subjects were instructed to (1) listen to the excerpt, (2) move to the music, changing movement when the music changed, (3) listen again, refer to the card if necessary, and tell what change occurred.

Written and Nonwritten Pitch, Duration, and Loudness scores were derived from the four measures. Reliability estimates computed from Main Study data for the derived scores were: Written Pitch, .73; Written Duration, .66; Written Loudness, .59; Nonwritten Pitch, .65; Nonwritten Duration, .55; Nonwritten Loudness, .35.

Procedures

Members of the research team proposed items for each section of the Battery of Musical Concept Measures. These items were reviewed, revised, and accepted or rejected before further validation of accepted items by a panel of judges. Several test-and-revise cycles prior to the Pilot Study involved trial administrations with various groups of subjects from the University Park area and two other Pennsylvania school districts. Items were examined according to item analysis criteria described in the study; items not functioning effectively were either revised or deleted.

The Battery was administered in the Pilot Study to two classrooms (56 subjects) from the Washington County, Maryland, schools. On the basis of item analysis, further revisions were made before the Battery was administered to the Main Study sample. In the Main Study the Written Measures (Listening and Verbal) were administered to sixteen entire classrooms of fourth-grade pupils; the Nonwritten Measures (Manipulative and Overt) were administered individually to pupils randomly selected from these classrooms. The data obtained were statistically analyzed to yield reliability estimates, correlation matrices, item analyses, and other statistics.

Reading achievement and IQ scores, for the segment of the sample with available data on California Mental Maturity and California Achievement Tests, were correlated with the scores from the Battery of Musical Concept Measures. Although correlations with the Written Measures were higher than with the Nonwritten Measures, all correlations were significant.

Subjects

Four hundred twenty-nine pupils from sixteen schools in twelve Pennsylvania school districts were tested on the group measures, in addition to the various trial-administration and Pilot Study samples. Subjects used in all trials totaled 601 elementary school children. From the sample of 429, 214 were randomly selected as subjects for the individual measures. The twelve school districts represented a large geographical area and various socioeconomic strata.

Conclusions and Implications

The Battery of Musical Concept Measures, a pioneer effort in the particular area of musical measurement involved, represents the development of a technique for identifying children's concepts of pitch, duration, and loudness. The Battery has functioned with considerable effectiveness during the present study in identifying fourth-grade pupils' musical concepts. Before the measures are acceptable for practical use, however, certain revisions and refinements should be made, and the measures standardized. At present, the measures are adequate for research purposes.

The Written Measures (Listening and Verbal), with their high reliabilities and efficiency of administration, show exceptional potential as practical measures. The Nonwritten Measures (Manipulative and Overt) are more time-consuming than the Written Measures in terms of the individual administration technique involved, and required trained personnel; nevertheless, these measures appear to possess particular possibilities for use with children having reading and/or language problems, or with younger children in the pre-reading stage.

In a number of cases it appeared that the subject possessed the concept being measured but exhibited confusion regarding the appropriate label (terminology). This suggests a need for increased emphasis on teaching labels. A test-teach-retest procedure with the Battery should yield further information concerning this aspect of the study, and also should furnish additional evidence on the validity of the measures.

In the development of the Battery of Musical Concept Measures, considerable effort was devoted to devising loudness and duration items of appropriate difficulty. The Manipulative Measure posed particular problems related to devising sufficiently difficult loudness items; it was necessary to approach a physiological level of discrimination to frame items with difficulty indices lower than .80. (It should be noted that high means or high difficulty indices reported on tests may indicate that the items devised were faulty, or that the subjects possessed the concept being measured.) The means and difficulty indices for Nonwritten Duration and Loudness were quite high, manipulative loudness being highest even after numerous efforts to devise more difficult items. Williams (56) and Riley and McKee (42) have indicated from their observations that the concept of loudness seems to develop early; Riley and McKee particularly noted young children's difficulty in learning pitch discrimination. Petzold (39) concluded from his study that

children perceive rhythm more accurately than melody, with later performance plateaus occurring for pitch-related tasks. In summarizing experimental studies, Valentine (53) indicated that non-trained listeners notice rhythm more than melody or harmony. These conclusions support indications from the present study. The authors believe that most children possess a more highly developed concept of loudness, and possibly of duration, than of pitch, as expressed on the Manipulative and Overt Measures.

In the Overt Measure, subjects evidenced change of overt movement in response to musical changes of duration and loudness; however, the majority of subjects failed to show such changes when pitch was the predominant change in the music. In correlating the overt responses with the subject's oral responses to the same musical stimuli, no significant relationship was found, indicating that subjects who could show the change might be unable to label the change, and that subjects who could orally indicate the change might not be able to demonstrate the change by overt movement.

Although the data are not conclusive, it appears that some children confuse the three terms high, loud, and fast, and the three terms low, soft, and slow. This may be due to the frequent association in music of these groups of characteristics, or to a general confusion of labels.

Most of the subjects appeared to enjoy participating in the study. Some children who were not selected from the classroom samples for the individual measurement expressed disappointment at being bypassed. Conversely, a few inhibited or self-conscious children seemed to find it difficult to move freely in the Overt Measure activity. Some of the apparent pupil enthusiasm may have resulted from the change in routine introduced by the testing, but a majority of the subjects evidently enjoyed the variety of musical experiences represented in the measures.

Recommendations for Future Research

Among the many truisms on the subject of research, there is one to the effect that "research begets research." The present study has suggested a number of avenues for future research. The authors propose to extend the study in the following ways, when funds are available:

1. The present chapter refers to a test-teach-retest procedure to be undertaken for the purpose of construct validation. Application has been made for funds to cover this procedure.

2. The dimension of timbre should be incorporated in both Written and Nonwritten Measures. This dimension was excluded from the present study after careful consideration. It is believed that this was a wise decision. The dimension is extremely difficult to place within a verbal mode, as a search of the literature describing instrumental and vocal timbre attests. There are few verbal descriptions of the tone quality of various instruments, and those that exist are traditional (in some cases almost whimsical) and certainly subjective.

3. The subject of children's growth in concept development (originally part of the present study in its unrevised form) should be explored through administering the measure to samples of subjects at different age or grade levels. (Grades 1, 3, and 5 had been noted in the original proposal prior to its revision.)

4. The Written Measures should develop even higher reliability through revision of certain items and/or lengthening of the measures. However, the authors believe that these measures, which in their present form may be administered in two thirty-minute sessions (time approximate), are near the tolerance limit of fourth-grade pupils' attention span. Item revision, rather than lengthening of measures, seems to be the better approach.

5. The Written Measures should be examined with the objective of adaptation for use with pre-school, first-grade, and second-grade pupils. This would make possible a longitudinal study of the development of musical concepts.

6. The Nonwritten Measures are believed to possess potential for further development. The authors feel that these measures (Overt and Manipulative) may prove of particular value with children handicapped verbally or visually, children who have problems speaking or reading, and possibly, children with problems relating to emotions or attitudes. Even during the testing that was part of the measures' development, the research team tentatively identified a few children who appeared to have certain of the problems mentioned. (These measures, however, when refined, should be effective with "normal" children also.) Of course, it is true that individual observation of a child, such as takes place during the administration of the Nonwritten Measures, is informative and revealing in itself.

In addition to the preceding proposed research, studies are needed in the following areas:

1. Studies are needed of children's reactions to various types of musical input constituting the daily media-derived environment as this affects concept development.

2. Further exploration is needed of the present research team's observation that the pitch stimulus in itself does not cause overt reaction manifested as physical movement.

3. Research in depth is indicated on the subject of how children acquire the observed semantic confusion as to musical terms or labels.

4. Studies of musical concepts based on children's free verbal reaction to musical excerpts could be revealing as to descriptions of music made in the characteristic vocabulary of children.

5. The study of children's overt reaction to musical stimulus for the purpose of demonstrating possession of musical concepts might be implemented by use of a videotaped or filmed record.

REFERENCES

1. Bartholomew, Wilmer T. Acoustics of Music. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1942. 242p.

2. Berelson, Bernard and Steiner, Gary. Human Behavior: An Inventory of Scientific Findings. New York: Harcourt, Brace and World, Inc. 1964. 712p.

3. Bergethon, Bjornar and Boardman, Eunice. Musical Growth in the Elementary School. New York: Holt, Rinehart and Winston, Inc. 1963. 276p.

4. Bernstein, Martin. An Introduction to Music. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1941. 396p.

5. Boekelheide, Viola E. Some Techniques of Assessing Certain Basic Music Listening Skills of Eight and Nine Year Olds. Unpublished doctoral dissertation, Stanford University. 1960.

6. Bond, Marjorie. "Rhythmic Perception and Gross Motor Performance." Research Quarterly. XXX, October 1959. pp.259-65.

7. Bridges, Virginia A. An Exploratory Study of the Harmonic Discrimination Ability of Children in Kindergarten through Grade Three in Two Selected Schools. Unpublished doctoral dissertation, Ohio State University. 1965.

8. Campbell, Donald T. "Recommendations for APA Test Standards Regarding Construct, Trait, or Discriminant Validity." The American Psychologist. XV, 1960. pp. 546-53.

9. Campbell, D. T. and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix." Psychological Bulletin. LVI, 1959. pp. 81-105.

10. Chan, Adrian and Travers, Robert M. W. "The Effect on Retention of Labeling Visual Displays." American Educational Research Journal. 3:1, January 1966. pp. 55-67.

11. Cronbach, Lee J. Essentials of Psychological Testing. Second Edition. New York: Harper and Row, Publishers. 1960. 650p.

12. Cronbach, Lee J. and Meehl, Paul E. "Construct Validity in Psychological Tests." Psychological Bulletin. LII, May 1955. pp. 281-302.

13. Culver, Charles A. Musical Acoustics. Second Edition. Philadelphia: The Blackiston Company. 1947. 215p.

14. Deihl, Ned C. Certain Relationships Among Concept Development, Listening Achievement, Musicality, and the Quantification of Musical Performance Experience. Unpublished doctoral dissertation, The Pennsylvania State University. 1963.

15. Ernst, Karl D. and Gary, Charles L. (eds.). Music in General Education. Washington: Music Educators National Conference. 1965. 223p.

16. Fleming, William and Veinus, Abraham. Understanding Music. New York: Henry Holt and Company. 1958. 452p.

17. Franklin, Erik. Tonality as a Basis for the Study of Musical Talent. Lund: Gumperts Forlag-Goteberg. 1956. Reported in Pflederer, Marilyn. "How Children Conceptually Organize Sounds." Council for Research in Music Education. Bulletin No. 7, Spring 1966. pp. 1-12.

18. Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill Book Company, Inc. 1956. 565p.

19. Harriman, Philip L. (ed.). The New Dictionary of Psychology. New York: Philosophical Library. 1947. 364p.

20. Harris, Dale B. Children's Drawings as Measures of Intellectual Maturity. New York: Harcourt, Brace and World, Inc. 1963. 367p.

21. Hartshorn, William. "The Teaching of Music." Music Educators Journal. 50:2, November-December 1963. pp. 28-9.

22. Hebb, D. O. Organization of Behavior. New York: Science Editions, Inc. 1961. 335p.

23. Heidbreder, E. F. "A Study of the Evolution of Concepts." Psychological Bulletin. Vol. 31, 1934. p. 673.

24. Hermann, Edward J. Supervising Music in the Elementary School. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1965. 210p.

25. Hoffer, Charles R. Teaching Music in the Secondary Schools. Belmont, California: Wadsworth Publishing Company, Inc. 1965. 463p.

26. Hunt, Earl B. Concept Learning. New York: John Wiley and Sons. 1962. 286p.

27. Il'ina, C. A. "On the Formation of Musical Conceptions in Preschool Children." Psychological Abstracts. XXXIV, 7332, 1960. p. 710.

28. Jeffrey, Wendell E. "Variables in Early Discrimination Learning: II. Mode of Response and Stimulus Difference in the Discrimination of Tonal Frequencies." Child Development. XXIX:4, December 1958. pp. 531-38.

29. LaRue, Jan. "On Style Analysis." pp. 139-41 in Perspectives in Music Education: Source Book III. Washington: Music Educators National Conference. 1966. 575p.

30. Leonhard, Charles and House, Robert W. Foundations and Principles of Music Education. New York: McGraw-Hill Book Company, Inc. 1959. 375p.

31. Lundin, Robert W. An Objective Psychology of Music. New York: The Ronald Press Company. 1953. 303p.

32. McDonald, Frederick J. Educational Psychology. San Francisco: Wadsworth Publishing Company, Inc. 1959. 748p.

33. McKee, John P. and Riley, Donald A. "Auditory Transposition in Six-Year-Old Children." Child Development. XXXIII, 1962. pp. 469-76.

34. McNemar, Quinn. Psychological Studies. Third Edition. New York: John Wiley and Sons, Inc. 1962. 451p.

35. Mainwaring, James. "Experiments on the Analysis of Cognitive Processes Involved in Musical Ability and in Musical Education." The British Journal of Educational Psychology. I, January 1931. pp. 180-203.

36. Moles, Abraham. Information Theory and Esthetic Perception. Translated by Joel E. Cohen. Urbana, Illinois: University of Illinois Press. 1966. 217p.

37. Monsour, Sally and Perry, Margaret. A Junior High School Music Handbook. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1963. 135p.

38. Palisca, Claude V. Music in Our Schools: A Search for Improvement. Report of the Yale Seminar on Music Education. Washington: U. S. Department of Health, Education, and Welfare, Office of Education, OE-33033, Bulletin 1964, No. 28. 61p.

39. Petzold, Robert G. Auditory Perception of Musical Sounds by Children in the First Six Grades. Unpublished report, Cooperative Research Project No. 1051, Office of Education, U. S. Department of Health, Education, and Welfare. 1966. 277p.

40. Pflederer, Marilyn. The Responses of Children to Musical Tasks Embodying Piaget's Principle of Conservation. Unpublished doctoral dissertation, University of Illinois. 1963.

41. Pflederer, Marilyn. "Responses of Children to Musical Tasks Embodying Piaget's Principle of Conservation." Journal of Research in Music Education. XII, Winter 1964. pp. 251-68.

42. Riley, Donald A. and McKee, John P. "Pitch and Loudness Transposition in Children and Adults." Child Development. XXXIV, 1963. pp. 471-82.

43. Royer, Fred. The Formation of Concepts with Auditory Stimuli. Unpublished doctoral dissertation, The Pennsylvania State University. 1956.

44. Russell, D. H. Children's Thinking. Waltham, Massachusetts: Blaisdell Publishing Company, A Division of Ginn and Company. 1956. 449p.

45. Russell, D. H. "Concepts." p. 323 in Encyclopedia of Educational Research. New York: The MacMillan Company. 1958.

46. Schoen, Max. Psychology of Music. New York: The Ronald Press Company. 1940. 258p.

47. Sheehy, Emma D. Children Discover Music and Dance. New York: Henry Holt and Company, Inc. 1959. 282p.

48. Simon, C. " Perceptual and Conceptual Functioning in the Musical Skills of Children Aged Six through Nine Years Old." Unpublished research reported in Pflederer, Marilyn. "How Children Conceptually Organize Sounds." Council for Research in Music Education. Bulletin No. 7, Spring 1966. pp. 1-12.

49. Simpson, Shirley E. "Development and Validation of an Objective Measure of Locomotor Response to Auditory Rhythmic Stimuli." Research Quarterly. XXIX, October 1958. pp. 342-48.

50. Smoke, K. L. "An Objective Study of Concept Formation." Psychological Monographs. 42:4, 1932. 46p.

51. Thompson, Oscar. "The Language of Music." from How to Understand Music. Fawcett Publications. 1936. pp. 22-5 in Ostransky, Leroy. Perspectives on Music. Englewood Cliffs, N. J.: Prentice-Hall, Inc. 1965. 430p.

52. Thorndike, Robert L. and Hagen, Elizabeth. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc. 1961. 602p.

53. Valentine, C. W. The Experimental Psychology of Beauty. London: Methuen and Company, Ltd. 1962. 280p.

54. Vinacke, W. Edgar. "Concept Formation in Children of School Ages." pp. 294-5 in Seidman, Jerome M. (ed.). The Child. New York: Rinehart and Company, Inc. 1958. 674p.

55. Ward, T. W. The Experimental Testing of a Concept of Perception with Implications for Music Education. Unpublished doctoral dissertation, University of Florida. 1956.

56. Williams, Harold; Sievers, Clement; and Hattwick, Melvin. "The Measurement of Musical Development." University of Iowa Studies in Child Welfare. VII:1, 1932. 191p.

57. Woodruff, Asahel D. "Concept Teaching in Music." pp. 219-223 in Perspectives in Music Education: Source Book III. Washington: Music Educators National Conference. 1966. 575p.

58. Woodruff, Asahel D. "The Nature and Elements of the Cognitive Approach to Education." Paper, 1964. Cited in A Proposal for the Revision of the Pre-Service Component of a Program of Teacher Education. Washington: The American Association of Colleges for Teacher Education. 1964. 66p.

59. Woodworth, Robert S. and Schlosberg, Harold. Experimental Psychology. Revised Edition. New York: Holt, Rinehart and Winston, Inc. 1963. 948p.

60. Zuckerkandl, Victor. The Sense of Music. Princeton, N. J.: The University Press. 1959. 246p.

Appendix A: Statistical Tables from Prepilot Studies

TABLE LI

VERBAL MEASURE ITEM ANALYSIS*
ESTELLA TRIAL
N = 31

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	P	.37	.75	.56	.39
2	P	.12	.75	.42	.63
3	D	.00	.62	.24	.77
4	L	.50	.50	.50	.00
5	D	.75	1.00	.90	.57
6	D	.12	.75	.42	.63
7	L	.00	.62	.24	.77
8	D	.37	.88	.64	.54
9	P	.00	.62	.24	.77
10	P	.12	.50	.29	.46
11	L	.75	1.00	.90	.57
12	P	.12	.88	.50	.73
13	P	.37	.88	.64	.54
14	D	.20	.62	.40	.44
15	P	.00	.62	.24	.77
16	P	.12	.88	.50	.73
17	D	.00	.75	.30	.82
18	L	.12	.75	.42	.63
19	P	.00	.12	.05	.43
20	P	.00	.62	.24	.77
21	D	.00	.00	.00	.00
22	D	.12	.62	.35	.54
23	P	.00	.62	.24	.77

*Fan table

TABLE LII

VERBAL MEASURE ITEM ANALYSIS*
CENTRAL DAUPHIN TRIAL
N = 24

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	P	.17	.83	.50	.65
2	P	.17	.83	.50	.65
3	D	.17	.83	.50	.65
4	L	.17	.50	.33	.37
5	D	.83	.83	.83	.00
6	D	.66	.83	.75	.22
7	L	.17	.50	.33	.37
8	D	.50	1.00	.81	.72
9	P	.17	.83	.50	.65
10	P	.00	.83	.34	.85
11	L	.66	1.00	.87	.64
12	P	.17	.66	.40	.50
13	P	.17	.66	.40	.50
14	D	.17	.83	.50	.65
15	P	.00	.66	.26	.79
16	P	.50	.66	.58	.17
17	D	.50	1.00	.81	.72
18	L	.33	.66	.49	.33
19	P	.17	.33	.25	.21
20	P	.00	.50	.19	.72
21	D	.17	.00	.07	.50
22	D	.17	1.00	.66	.85
23	P	.00	.17	.07	.50

*Fan table

TABLE LIII

VERBAL MEASURE ITEM ANALYSIS*
CENTRAL DAUPHIN AND
ESTELLA COMBINED
N = 55

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	P	.27	.67	.47	.40
2	P	.13	.67	.38	.56
3	D	.07	.67	.34	.65
4	L	.33	.60	.46	.28
5	D	.74	.93	.84	.32
6	D	.33	.74	.54	.41
7	L	.13	.53	.31	.45
8	D	.40	.87	.65	.51
9	P	.07	.74	.37	.69
10	P	.07	.67	.34	.65
11	L	.74	1.00	.90	.58
12	P	.13	.67	.38	.56
13	P	.27	.80	.54	.53
14	D	.27	.60	.43	.34
15	P	.00	.67	.26	.79
16	P	.33	.74	.54	.41
17	D	.27	.93	.63	.68
18	L	.13	.80	.45	.66
19	P	.13	.20	.16	.12
20	P	.27	.27	.27	.00
21	D	.20	.00	.08	-.53
22	D	.13	.80	.45	.66
23	P	.00	.20	.08	.53

*Fan table

TABLE LIV

LISTENING ITEMS, PERCENTAGES CORRECT
UNIVERSITY GROUPS TRIALS

Item number	Dimension measured	Per cent of Correct Responses		
		Mu Ed 574	Mu Ed 86	BOC Chorus
1	L	50.0	88.9	92.9
2	D	58.3	88.9	100.0
3	P	58.3	55.6	85.7
4	P	58.3	61.1	78.6
5	D	66.7	88.9	96.4
6	L	83.3	88.9	96.4
7	P	66.7	94.4	96.4
8	D	50.0	72.2	92.9
9	L	83.3	94.4	100.0
10	P	58.3	55.6	53.7
11	P + L	83.3	33.3	60.7
12	D + L	83.3	77.8	89.3
13	P + L	91.7	88.9	100.0
14	P + L	91.7	88.9	100.0
15	L	100.0	66.7	82.1
16	D	91.7	61.1	71.4
17	D	75.0	83.3	85.7
18	L	66.7	77.8	71.4
19	P	91.7	94.4	89.3
20	P	100.0	100.0	92.9

TABLE LV

LISTENING MEASURE ITEM ANALYSIS*
UNIVERSITY AREA TRIAL
N = 20

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.60	1.00	.85	.67
2	D	.60	.60	.60	.00
3	P	.20	.80	.50	.59
4	P	.20	.80	.50	.59
5	D	.80	1.00	.92	.53
6	L	.40	1.00	.77	.76
7	P	.60	1.00	.85	.67
8	D	.60	1.00	.85	.67
9	L	.80	1.00	.92	.53
10	P	.60	1.00	.85	.67
11	P + L	.20	.60	.39	.42
12	D + L	.60	1.00	.85	.67
13	P + L	.00	1.00	.50	.93
14	D + L	.20	.80	.50	.59
15	D + L	.20	1.00	.67	.85
16	P + L	.60	1.00	.85	.67
17	P	.80	.80	.80	.00
18	D	.80	1.00	.92	.53
19	P	.60	1.00	.85	.67
20	L	.60	.80	.70	.24
21	D	.40	.60	.50	.20
22	D	.40	.20	.30	-.24
23	P	1.00	1.00	1.00	.00
24	L	.80	1.00	.92	.53

*Fan table

TABLE LVI

LISTENING MEASURE ITEM ANALYSIS*
ESTELLA TRIAL
N = 31

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.00	.75	.30	.82
2	D	.37	.50	.43	.13
3	P	.00	.12	.05	.43
4	P	.50	.25	.37	-.27
5	D	.25	1.00	.70	.82
6	L	.00	.62	.24	.77
7	P	.00	1.00	.50	.93
8	D	.12	.50	.29	.44
9	L	.12	.75	.42	.63
10	P	.37	.87	.64	.53
11	P + L	.25	.50	.37	.27
12	D + L	.50	.87	.70	.43
13	P + L	.50	.87	.70	.43
14	D + L	.12	.62	.35	.54
15	D + L	.12	.75	.42	.63
16	P + L	.37	.75	.56	.39
17	P	.00	.37	.14	.66
18	D	.00	.75	.30	.82
19	P	.00	.62	.24	.77
20	L	.00	.75	.30	.82
21	D	.00	.62	.24	.77
22	D	.12	.25	.18	.20
23	P	.12	1.00	.63	.87
24	L	.12	.75	.42	.63

*Fan table

TABLE LVII

MEANS AND STANDARD DEVIATIONS
FOR THE LISTENING MEASURE
PREPILOT TRIALS

Subjects	Number of subjects	Mean	Standard deviation
University area children	20	17.4	3.38
Estella	31	9.5	4.84
Central Dauphin	24	11.96	4.46
Combined (Central Dauphin and Estella)	55	10.59	4.79

TABLE LVIII

LISTENING MEASURE ITEM ANALYSIS*
CENTRAL DAUPHIN AND
ESTELLA COMBINED
N = 55

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.07	.87	.45	.77
2	D	.47	.67	.57	.21
3	P	.07	.33	.18	.40
4	P	.27	.60	.43	.34
5	D	.47	1.00	.80	.74
6	L	.00	.67	.26	.79
7	P	.07	1.00	.59	.89
8	D	.13	.60	.35	.51
9	L	.27	.80	.54	.53
10	P	.20	.80	.50	.59
11	P + L	.47	.47	.47	.00
12	D + L	.67	.80	.74	.16
13	P + L	.60	.47	.54	-.13
14	D + L	.33	.74	.54	.41
15	D + L	.20	.80	.50	.59
16	P + L	.33	.74	.54	.41
17	P	.00	.40	.15	.67
18	D	.20	.80	.50	.59
19	P	.00	.80	.33	.84
20	L	.13	.74	.42	.61
21	D	.00	.47	.18	.71
22	D	.07	.33	.18	.40
23	P	.13	.87	.50	.72
24	L	.13	.93	.55	.77

*Fan table

TABLE LIX

MANIPULATIVE MEASURE ITEM ANALYSIS*
ESTELLA TRIAL
N = 31

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	L	.75	1.00	.90	.57
2	L	.50	1.00	.81	.72
3	L	1.00	1.00	1.00	.00
4	L	1.00	1.00	1.00	.00
5	D	.25	.87	.58	.62
6	D	.25	1.00	.70	.82
7	P	.12	.75	.42	.63
8	P	.00	.75	.30	.82
9	P	.00	.62	.24	.77
10	P	.00	.62	.24	.77
11	D	.50	1.00	.81	.72
12	D	.62	1.00	.85	.66
13	P	.12	.87	.49	.63
14	P	.00	1.00	.50	.93
15	D	1.00	1.00	1.00	.00
16	D	.87	1.00	.94	.45
17	L	1.00	1.00	1.00	.00
18	L	1.00	1.00	1.00	.00

*Fan table

TABLE LX

MANIPULATIVE MEASURE ITEM ANALYSIS*
BELLEFONTE TRIAL
N = 20

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1	D	.17	.83	.50	.65
2	D	.33	.83	.59	.51
3	L	.83	1.00	.93	.50
4	D	.50	1.00	.81	.72
5	D	.50	1.00	.81	.72
6	P	.17	.83	.50	.65
7	P	.17	.67	.41	.51
8	P	.17	.83	.50	.65
9	P	.17	.33	.25	.21
10	L	1.00	1.00	1.00	.00
11	L	.67	1.00	.87	.63
12	P	.00	.50	.19	.72
13	L	1.00	1.00	1.00	.00
14	D	1.00	1.00	1.00	.00
15	L	1.00	.83	.93	-.50
16	P	.00	.33	.13	.63
17	L	1.00	1.00	1.00	.00
18	D	.00	.33	.13	.63

*Fan table

TABLE LXI

OVERT MEASURE ITEM ANALYSIS*
ESTELLA TRIAL
N = 31

Item number	Dimension measured	Proportion of lowest 26% passing item	Proportion of highest 26% passing item	Item difficulty estimate	Discrimination index
1a	L	.00	1.00	.50	.93
1b	L	.00	1.00	.50	.93
2a	D	.25	.88	.58	.63
2b	D	.25	.88	.56	.63
3a	D	.25	1.00	.70	.82
3b	D	.25	1.00	.70	.82
4a	D	.25	1.00	.70	.82
4b	D	.12	1.00	.63	.87
5a	P	.12	1.00	.63	.87
5b	P	.12	.88	.50	.73
6a	L	.00	.88	.37	.87
6b	L	.00	.88	.37	.87
7a	D	.25	1.00	.70	.82
7b	D	.25	1.00	.70	.82
8a	P	.37	1.00	.75	.78
8b	P	.37	1.00	.75	.78
9a	L	.12	1.00	.63	.87
9b	L	.12	1.00	.63	.87
10a	P	.12	.12	.12	.00
10b	P	.12	.00	.05	-.43
11a	D	.25	1.00	.70	.82
11b	D	.25	1.00	.70	.82

*Fan table

Appendix B: Measures Used Prior to Main Study

Verbal Measure, Form 1

PICK THE BEST ANSWER

1. As a melody goes up, the tones must become:

   a. softer
   b. louder
   c. higher
   d. lower
   e. faster

2. As a melody goes down, the tones must become:

   a. faster
   b. higher
   c. louder
   d. softer
   e. lower

3. A short note could be:

   a. loud
   b. soft
   c. high
   d. low
   e. any of these

4. When music gets louder we hear:

   a. more sound
   b. less sound
   c. higher sound
   d. lower sound
   e. slower sound

5. Fast music is most like someone:

   a. walking
   b. running
   c. creeping
   d. hopping
   e. standing

6. If you played a song faster and louder the second time,

   a. it would take less time than the first time
   b. it would take more time than the first time
   c. it would be higher than the first time
   d. it would be lower than the first time
   e. any of these

7. When music gets softer, we hear:

   a. higher sound
   b. lower sound
   c. faster sound
   d. less sound
   e. more sound

8. Slow music is most like:

   a. running
   b. racing
   c. creeping
   d. galloping
   e. whirling

9. High music is most like:

   a. thunder
   b. father's voice
   c. singing birds
   d. rumbling trucks
   e. roaring cannons

10. Low music is most like:

   a. doorbell
   b. thunder
   c. chirping birds
   d. mother's voice
   e. telephone

11. As a band in a parade gets near to you,

   a. the music seems louder
   b. the music seems softer
   c. the music seems higher
   d. the music seems lower
   e. the notes seem faster

12. If you sing out of tune, this means:

   a. you sing too loudly
   b. you sing too softly
   c. you sing too quickly
   d. you sing ahead or behind the class
   e. you sing higher or lower than the right pitches

13. Which instrument can sound several notes at the same time?

   a. drum
   b. piano
   c. trumpet
   d. flute
   e. voice

14. If you wrote a song for marching, it would be:

a. faster than a lullaby
b. slower than a lullaby
c. soft and slow
d. loud and slow
e. high and slow

15. A high note could be:

a. loud
b. soft
c. long
d. short
e. any of these

16. The opposite of a high tone is:

a. loud tone
b. soft tone
c. low tone
d. long tone
e. short tone

17. In marching we step to the:

a. loudness of the music
b. softness of the music
c. the beat of the music
d. the lowness of the music
e. any of these

18. A loud note could be:

    a. high
    b. low
    c. long
    d. short
    e. any of these

19. Pitch tells us:

    a. how soft the music is
    b. how fast the music is
    c. how high the music is
    d. how light the music is
    e. all of these

20. If you played a song lower and louder the second time, it would:

    a. take more time than the first time
    b. take less time than the first time
    c. take more notes the second time
    d. have different notes the second time
    e. have the same notes the second time

21. Two sounds of equal length must be:

    a. equally soft
    b. equally short
    c. equally loud
    d. equally high
    e. any of these

22. If you were to sing a song slower and higher the second time,

    a. it would be softer
    b. it would be louder
    c. it would take more time to sing
    d. it would take less time to sing
    e. it would take the same amount of time

23. Playing a song softer and lower the second time would mean:

    a. you would finish it sooner the second time
    b. it would sound like a different tune the second time
    c. there would be no change in the tune
    d. it would be faster the second time
    e. it would be slower the second time

Name _____
Hagerstown

Verbal Measure, Form 2

Choose the best answer to complete each statement. Fill in the answer space before the answer you think is best. Choose only one answer for each number.

Sample

If you sing cut of tune, this means

- ☐ you sing too quickly.
- ☐ you sing too softly.
- ☐ you sing too loudly.
- ☐ you sing higher or lower than the right pitches.
- ☐ you sing ahead of or behind the class.

1. As a band in a parade gets near to you,
   - ☐ the music seems louder.
   - ☐ the music seems higher.
   - ☐ the music seems softer.
   - ☐ the music seems lower.
   - ☐ the notes seem faster.
2. Low music is most like
   - ☐ a doorbell.
   - ☐ thunder.
   - ☐ chirping birds.
   - ☐ mother's voice.
   - ☐ the telephone.
3. If you were to sing a song slower and higher the second time,
   - ☐ it would be softer.
   - ☐ it would be louder.
   - ☐ it would take more time to sing.
   - ☐ it would take less time to sing.
   - ☐ it would take the same amount of time.

4. If you wrote a song for marching, it would be

- ☐ faster than a lullaby.
- ☐ slower than a lullaby.
- ☐ soft and slow.
- ☐ loud and slow.
- ☐ high and slow.

5. A short note could be

- ☐ loud.
- ☐ soft.
- ☐ high.
- ☐ low.
- ☐ any of these.

6. When music gets softer, we hear

- ☐ higher sound.
- ☐ lower sound.
- ☐ faster sound.
- ☐ less sound.
- ☐ more sound.

7. If you played a song higher and softer the second time, it would

- ☐ sound bigger the second time.
- ☐ sound smaller the second time.
- ☐ take less time the second time.
- ☐ take more time the second time.
- ☐ use more notes the second time.

8. If you played a song faster and louder the second time,

- ☐ it would take less time than the first time.
- ☐ it would take more time than the first time.
- ☐ it would be higher than the first time.
- ☐ it would be lower than the first time.
- ☐ any of these.

9. The opposite of a high tone is a

- ☐ loud tone.
- ☐ soft tone.
- ☐ low tone.
- ☐ long tone.
- ☐ short tone.

10. In marching, we step to the

- ☐ loudness of the music.
- ☐ softness of the music.
- ☐ the beat of the music.
- ☐ the lowness of the music.
- ☐ any of these.

11. When music gets louder, we hear

- ☐ more sound.
- ☐ less sound.
- ☐ higher sound.
- ☐ lower sound.
- ☐ faster sound.

12. A loud note could be

- ☐ high.
- ☐ low.
- ☐ long.
- ☐ short.
- ☐ any of these.

13. As a melody goes up, the tones must become

- ☐ softer.
- ☐ louder.
- ☐ higher.
- ☐ lower.
- ☐ faster.

14. Playing a song louder and lower the second time would mean

- ☐ you would finish it sooner the second time.
- ☐ there would be less sound the second time.
- ☐ there would be more sound the second time.
- ☐ it would be faster the second time.
- ☐ it would be slower the second time.

15. High music is most like

- ☐ thunder.
- ☐ croaking frogs.
- ☐ singing birds.
- ☐ a rumbling truck.
- ☐ roaring cannons.

16. Slow music is most like

- ☐ running.
- ☐ skipping.
- ☐ creeping.
- ☐ galloping.
- ☐ whirling.

17. Which instrument can sound several notes at the same time?

- ☐ drum
- ☐ piano
- ☐ trumpet
- ☐ flute
- ☐ voice

18. A high note could be

- ☐ loud.
- ☐ soft.
- ☐ long.
- ☐ short.
- ☐ any of these.

Musical Sources of Items and Correct Responses
Listening Measure, Form 1

	Musical Source	Correct Response
1.	Sullivan: Iolanthe, Overture	louder
2.	Sullivan: Pinafore, Overture	softer
3.	Sullivan: Pinafore, Overture	slower
4.	Sullivan: Pinafore, Overture	higher
5.	Bizet: Carmen, Habanera	lower
6.	Lecuona: Malaguena	faster
7.	Beethoven: Symphony No. 3, Third Movement	louder
8.	Bach-Stokowski: Toccata in D Minor	lower
9.	Tchaikowsky: Swan Lake, Czardas	faster
10.	Grofe: Grand Canyon Suite, Sunrise	higher
11.	Handel: Water Music, Andante	slower
12.	Handel: Fireworks, Lentement	softer
13.	Beethoven: Symphony No. 1, Third Movement	higher and louder
14.	Beethoven: Leonore Overture No. 2	slower and lower
15.	Balfe: The Bohemian Girl, Overture	slower and softer
16.	Beethoven: Leonore Overture No. 2	higher and louder
17.	Beethoven: Leonore Overture No. 2	higher and louder
18.	Brahms: Variations on a Theme by Haydn	softer
19.	Sullivan: Pirates of Penzance, Overture	faster
20.	Sullivan: Pirates of Penzance, Overture	slower
21.	Beethoven: String Quartet No. 3, Third Movement	higher
22.	Tchaikowsky: Swan Lake, Waltz	louder
23.	Tchaikowsky: Swan Lake, Little Swans	higher
24.	Beethoven: String Quartet No. 5, Third Movement	lower

Answer Sheet, Listening Measure, Form 1

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____
11. _____
12. _____
13. _____
14. _____
15. _____
16. _____
17. _____
18. _____
19. _____
20. _____
21. _____
22. _____
23. _____
24. _____

Musical Sources of Items and Correct Responses
Listening Measure, Form 2

	Musical Source	Correct Response
1.	Sullivan: Pinafore, Overture	softer
2.	Sullivan: Pinafore, Overture	slower
3.	Sullivan: Pinafore, Overture	higher
4.	Bizet: Carmen, Habanera	lower
5.	Lecuona: Malaguena	faster
6.	Beethoven: Symphony No. 3, Third Movement	louder
7.	Bach-Stokowki: Toccata in D Minor	lower
8.	Handel: Water Music, Andante	slower
9.	Handel: Water Music, Lentement	softer
10.	Beethoven: Leonore Overture No. 3	lower
11.	Beethoven: Symphony No. 1, Third Movement	higher and louder
12.	Tchaikowsky: Swan Lake, Czardas	faster and softer
13.	Beethoven: Leonore Overture No. 2	higher and louder
14.	Beethoven: Leonore Overture No. 2	higher and louder
15.	Brahms: Variations on a Theme by Haydn	softer
16.	Sullivan: Pirates of Penzance, Overture	faster
17.	Sullivan: Pirates of Penzance, Overture	slower
18.	Tchaikowsky: Swan Lake, Waltz	higher
19.	Tchaikowsky: Swan Lake, Little Swans	higher
20.	Beethoven: String Quartet No. 5, Third Movement	lower

Answer Sheet, Listening Measure, Form 2

We will hear some short musical selections. Each one will change in some way. Listen to the complete example and decide what is the most important change. Choose the correct answer on your paper, but listen again as the same example is repeated before marking your answer. Fill in the square beside the answer that you think is right.

Sample

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ higher

1.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ lower

2.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

3.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

4.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

5.

- ☐ faster
- ☐ slower
- ☐ softer
- ☐ higher
- ☐ lower

6.

- ☐ faster
- ☐ slower
- ☐ lower
- ☐ louder
- ☐ softer

7.

- ☐ louder
- ☐ softer
- ☐ lower
- ☐ higher
- ☐ faster

8.
- ☐ softer
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

9.
- ☐ softer
- ☐ louder
- ☐ higher
- ☐ lower
- ☐ slower

10.
- ☐ louder
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

11.
- ☐ higher and slower
- ☐ higher and louder
- ☐ lower and softer
- ☐ lower and louder
- ☐ louder and faster

12.
- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

13.
- ☐ higher and softer
- ☐ higher and louder
- ☐ lower and softer
- ☐ lower and louder
- ☐ lower and faster

14.
- ☐ higher and softer
- ☐ higher and louder
- ☐ higher and slower
- ☐ lower and slower
- ☐ lower and faster

The final group of examples is different. This time you must listen to two different examples. Decide how the second example is most different from the first. Fill in the square beside the word which tells this difference. These examples will not be repeated.

Sample

- ☐ louder
- ☐ softer
- ☐ higher
- ☐ slower
- ☐ faster

15.

- ☐ louder
- ☐ softer
- ☐ faster
- ☐ slower
- ☐ higher

16.

- ☐ higher
- ☐ softer
- ☐ louder
- ☐ slower
- ☐ faster

17.

- ☐ higher
- ☐ softer
- ☐ louder
- ☐ slower
- ☐ faster

18.

- ☐ faster
- ☐ slower
- ☐ softer
- ☐ louder
- ☐ lower

19.

- ☐ lower
- ☐ higher
- ☐ faster
- ☐ slower
- ☐ softer

20.

- ☐ lower
- ☐ higher
- ☐ louder
- ☐ faster
- ☐ slower

Musical Sources of Items and Correct Responses
Listening Measure, Form 3

	Musical Source	Correct Response
1.	Sullivan: Pinafore, Overture	softer
2.	Sullivan: Pinafore, Overture	slower
3.	Sullivan: Pinafore, Overture	higher
4.	Bizet: Carmen, Habanera	lower
5.	Lecuona: Malaguena	faster
6.	Beethoven: Symphony No. 3, Third Movement	louder
7.	Bach-Stokowski: Toccata in D Minor	lower
8.	Handel: Water Music, Andante	slower
9.	Handel: Fireworks, Lentement	softer
10.	Beethoven: Leonore Overture No. 2	lower
11.	Bach-Stokowski: Toccata in D Minor	louder and lower
12.	Holst: The Planets, Venus	slower and softer
13.	Beethoven: Symphony No. 1, Third Movement	higher and louder
14.	Holst: The Planets, Jupiter	faster and softer
15.	Tchaikowsky: Swan Lake, Czardas	faster and softer
16.	Beethoven: Leonore Overture No. 2	higher and louder
17.	Brahms: Variations on a Theme by Haydn	lower
18.	Brahms: Variations on a Theme by Haydn	faster
19.	Tchaikowsky: Swan Lake, Little Swans	higher
20.	Tchaikowsky: Swan Lake, Waltz	louder
21.	Sullivan: Pirates of Penzance, Overture	faster
22.	Sullivan: Pirates of Penzance, Overture	slower
23.	Brahms: Concerto for Violin and Cello	lower
24.	Ravel: Ma Mere L'Oye, Pavanne	softer

Answer Sheet, Listening Measure, Form 3

DIRECTIONS:

We will hear some short musical selections. Each one will change in some way. Listen to the complete example and decide what is the most important change. Choose the correct answer on your paper, but listen again as the same example is repeated before marking your answer. Then fill in the square beside the answer that you think is right.

SAMPLE A

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ higher

SAMPLE B

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

1.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ lower

2.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ softer

3.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

4.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

5.

- ☐ faster
- ☐ slower
- ☐ softer
- ☐ higher
- ☐ lower

6.

- ☐ louder
- ☐ softer
- ☐ faster
- ☐ slower
- ☐ lower

7.

- ☐ louder
- ☐ softer
- ☐ lower
- ☐ higher
- ☐ faster

8.

- ☐ softer
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

9.

- ☐ softer
- ☐ louder
- ☐ higher
- ☐ lower
- ☐ slower

10.

- ☐ louder
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

DIRECTIONS:

Look at your paper beginning with Number 11. In this part of the test there are two answers after each square. The music will do two things at once. It may get higher and slower or faster and louder. Be sure to choose the answer that has both parts right.

11.
- ☐ higher and slower
- ☐ higher and louder
- ☐ lower and softer
- ☐ lower and louder
- ☐ louder and faster

12.
- ☐ lower and softer
- ☐ lower and louder
- ☐ faster and softer
- ☐ slower and softer
- ☐ faster and louder

13.
- ☐ higher and slower
- ☐ higher and louder
- ☐ lower and softer
- ☐ softer and faster
- ☐ louder and faster

14.
- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

15.
- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

16.
- ☐ higher and softer
- ☐ higher and louder
- ☐ higher and slower
- ☐ lower and slower
- ☐ lower and faster

DIRECTIONS:

The final group of examples is different. This time you must listen to two different examples. Decide how the second example is most different from the first. Fill in the square beside the word which tells this difference. Listen carefully the first time, because these examples will not be repeated.

SAMPLE C

- ☐ faster
- ☐ slower
- ☐ higher
- ☐ lower
- ☐ louder

SAMPLE D

- ☐ louder
- ☐ softer
- ☐ higher
- ☐ slower
- ☐ faster

17.

- ☐ higher
- ☐ lower
- ☐ softer
- ☐ faster
- ☐ slower

18.

- ☐ lower
- ☐ louder
- ☐ softer
- ☐ faster
- ☐ slower

19.

- ☐ lower
- ☐ higher
- ☐ slower
- ☐ faster
- ☐ softer

20.

- ☐ louder
- ☐ slower
- ☐ faster
- ☐ higher
- ☐ lower

21.

- ☐ higher
- ☐ softer
- ☐ louder
- ☐ slower
- ☐ faster

22.

- ☐ higher
- ☐ softer
- ☐ louder
- ☐ slower
- ☐ faster

23.

- ☐ lower
- ☐ higher
- ☐ faster
- ☐ slower
- ☐ softer

24.

- ☐ lower
- ☐ higher
- ☐ faster
- ☐ louder
- ☐ softer

Musical Sources of Items and Correct Responses
Listening Measure, Form 4

	Musical Source	Correct Response
	Sample A Sullivan: Iolanthe, Overture	louder
	Sample B Beethoven: Leonore Overture No. 2	lower
1.	Sullivan: Pinafore, Overture	softer
2.	Prokofiev: Classical Symphony, First Movement	higher
3.	Bach-Stokowski: Toccata in D Minor	lower
4.	Beethoven: Symphony No. 3, Third Movement	louder
5.	Handel: Water Music, Hornpipe	slower
6.	Lecuona: Malaguena	faster
	Sample C Holst: The Planets, Jupiter	softer and faster
7.	Stravinsky: Petrouchka, Russian Dance	softer and slower
8.	Beethoven: Leonore Overture No. 2	higher and louder
9.	Tchaikowsky: Swan Lake, Czardas	softer and faster
10.	Prokofiev: Lieutenant Kije, Kije's Wedding	slower and lower
11.	Beethoven: Consecration of the House	higher and faster
12.	Lizst: Grand Galop Chromatique	louder and lower
	Sample D Holst: The Planets, Jupiter	lower
	Sample E Brahms: Variations on a Theme by Haydn	softer
13.	Brahms: Variations on a Theme by Haydn	faster
14.	Tchaikowsky: Swan Lake, Little Swans	higher
15.	Tchaikowsky: Swan Lake, Waltz	louder
16.	Glinka: Russlan and Ludmilla, Overture	lower
17.	Ravel: Ma Mere l'Oye, Pavanne	softer
18.	Tchaikowsky, Capriccio Italien	slower

Answer Sheet, Listening Measure, Form 4

DIRECTIONS:

We will hear some short musical examples. Each one will change in some way. Listen to the complete example and decide on the most important change. Read the answers on your paper, but before marking your answer listen again as the same example is repeated. After hearing the example the second time, fill in the square beside the answer you think is right. Be sure you fill in only one answer space for each example.

SAMPLE A

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ higher

SAMPLE B

- ☐ louder
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

1.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ lower

2.

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ higher
- ☐ lower

3.

- ☐ louder
- ☐ softer
- ☐ lower
- ☐ higher
- ☐ faster

4.

- ☐ louder
- ☐ softer
- ☐ faster
- ☐ slower
- ☐ lower

5.

- ☐ softer
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

6.

- ☐ faster
- ☐ slower
- ☐ softer
- ☐ higher
- ☐ lower

DIRECTIONS:

In this part there are two answers after each square. Listen as the music changes in two ways. For example, the music may get higher and slower--or faster and louder. Be sure to choose the answer that has both parts right, but fill in only <u>one</u> answer space for each example.

SAMPLE C

- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

7.

- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

8.

- ☐ higher and softer
- ☐ higher and louder
- ☐ higher and slower
- ☐ lower and louder
- ☐ lower and faster

9.

- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

10.

- ☐ higher and softer
- ☐ higher and louder
- ☐ higher and slower
- ☐ lower and slower
- ☐ lower and faster

11.

- ☐ lower and louder
- ☐ lower and faster
- ☐ higher and faster
- ☐ higher and slower
- ☐ louder and slower

12.

- ☐ lower and softer
- ☐ lower and louder
- ☐ higher and softer
- ☐ slower and softer
- ☐ higher and louder

DIRECTIONS:

In this group each example will have two different parts. Decide how the second part is most different from the first. Fill in the square beside the word which best describes this difference. Fill in only one answer space for each example. Listen carefully the first time because these examples will not be repeated.

SAMPLE D

- ☐ faster
- ☐ slower
- ☐ higher
- ☐ lower
- ☐ louder

SAMPLE E

- ☐ louder
- ☐ softer
- ☐ higher
- ☐ slower
- ☐ faster

13.

- ☐ lower
- ☐ louder
- ☐ softer
- ☐ faster
- ☐ slower

14.

- ☐ lower
- ☐ higher
- ☐ slower
- ☐ faster
- ☐ softer

15.

- ☐ louder
- ☐ slower
- ☐ faster
- ☐ higher
- ☐ lower

16.

- ☐ lower
- ☐ higher
- ☐ faster
- ☐ slower
- ☐ softer

17.

- ☐ lower
- ☐ higher
- ☐ faster
- ☐ louder
- ☐ softer

18.

- ☐ lower
- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer

Directions for Administration, Listening Measure, Pilot Study

Distribute answer sheets. Make sure that each child has pencil with eraser. Instruct children to write their names in the space provided on page 1.

START TAPE

*"At the top of your answer sheet there are some directions. Read them silently to yourself while I read them to you. /"We will hear some short musical examples. Each one will change in some way. Listen to the complete example and decide on the most important change. Read the answers on your paper, but before marking your answer, listen again as the same example is repeated. After hearing the example the second time, fill in the square beside the answer you think is right. Be sure you fill in only one answer space for each example."/

"Let's try a sample. Listen to Sample A and hear how it changes."

(Sample A)

"Look at your answer sheet and decide which answer best describes how the music changed. (Pause) Listen again before marking your answer to see whether you were correct."

(Sample A repeated)

"Mark your answer for Sample A."

STOP TAPE after "Sample A."

ASK: "Which word describes the most important change?"
ALLOW CHILDREN TO ANSWER (Correct answer is "louder").
SAY: Fill in the space beside "louder." If you had a different answer, erase it."
WHEN ALL HAVE CORRECTED SAMPLE A, SAY "Let's try another sample."

START TAPE

"Listen to Sample B."

(Sample B)

"Decide what you think is the biggest change. Look for that answer on the sheet, but listen again before marking your answer."

(Sample B repeated)

---

*Indented items are on the tape.

"Mark your answer for Sample B."

STOP TAPE after "Sample B."

AFTER CHILDREN HAVE MARKED ANSWERS, ASK, "What did you mark?" AFTER CHILDREN HAVE ANSWERED, "Lower is right." "Are there any questions about how to do this?" ANSWER ANY QUESTIONS ABOUT PROCEDURE, THEN SAY, "Let's begin with Number 1.

START TAPE

Items 1 to 6

"Turn to the directions at the top of page 2."

(Pause) /"In this part there are two answers after each square. Listen as the music changes in two ways. For example, the music may get higher and slower--or faster and louder. Be sure to choose the answer that has both parts right, but fill in only one answer space for each example."/

"Listen to a sample."

(Sample C)

"Decide the two most important ways the music changed and find the answer with both parts correct; but listen again before marking the answer."

(Sample C repeated)

"Mark your answer for Sample C."

STOP TAPE after "Sample C."

AFTER CHILDREN HAVE MARKED ANSWERS, ASK "What answer did you mark?" ALLOW CHILDREN TO ANSWER. (Correct answer is "faster and softer.") SAY "Be sure you have filled in the one space before the two words, faster and softer. Choose the one square with both correct answers, not two squares. Are there any questions?"

START TAPE after any questions have been answered.

Items 7 to 12

"Turn to the directions at the top of page 3."

(Pause) /"In this group each example will have two different parts. Decide how the second part is most different from the first. Fill in the square beside the word which best describes this difference. Fill in only one answer space for each example. Listen carefully

the first time because these examples will not be repeated."/ Let's try a sample. Fill in the answer space beside the word which tells how the second part was different from the first part."

(Sample D--two parts)

STOP TAPE after "Sample D."

ASK "What is the correct answer?" GET THE CORRECT ANSWER (lower) FROM THE CLASS. AFTER CHILDREN HAVE HAD TIME TO CORRECT THE ANSWER, "Let's try another sample."

START TAPE

(Sample E)

"Mark your answer for Sample E."

STOP TAPE after "Sample E."

GIVE CHILDREN TIME TO MARK AN ANSWER. ASK, "What answer do you have?" AFTER CHILDREN ANSWER (Correct answer is softer), ASK "Are there any questions?" AFTER ANSWERING QUESTIONS, "Let's go ahead with Number 13. Remember, there are two different parts for each example and these are played only once."

START TAPE

Items 13 to 18.

STOP TAPE at the end of Number 18, second part.

GIVE CHILDREN TIME TO COMPLETE THE ANSWER. ASK THEM TO FOLD PAGE BACK SO THAT THEIR NAMES ARE ON THE TOP SHEET. COLLECT THE PAPERS.

Answer Sheet, Listening and Overt Measures,
Final Forms, Preliminary Validation

Sample A _____

Sample B _____

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____

Sample C _____

7. _____
8. _____
9. _____
10. _____
11. _____
12. _____

Sample D _____

Sample E _____

13. _____
14. _____
15. _____
16. _____
17. _____
18. _____

Overt Sample _____

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____

Manipulative Measure, Form 1

1. Play 2 bells - which is higher
2. Play 3 bells - which is the middle pitch
3. Play 5 bells - arrange in order, play high to low
4. Play 4 glasses - arrange in order, play low to high

Experiment with bongos (play soft and loud)

5. Play higher drum soft
6. Play lower drum loud
7. Play drum fast and get slower
8. Play drum fast and soft
9. Play drum slow and loud

Experiment with orchestra bells (show loud and soft)

10. Play loudly a scale getting higher
11. Play softly a scale getting lower

Experiment with chord organ

12. Play short, long, and medium notes
13. Using black buttons, play the chord which is highest, play lowest
14. Using the 3 wood blocks, play the highest
15. Play 3 finger cymbals in order from low to high

Manipulative Measure, Form 2

NAME _____

ITEM	RESPONSE Right	Wrong	TESTING
DRUM 1. Play fast and softly			L
2. Play loudly and slowly			L
3. Play loudly and get softer			L
4. Play softly and get louder			L
5. Play fast and get slower			D
6. Play slowly and get faster			D
WATER GLASSES 7. Play from low to high			P
UKULELE 8. Play from high to low			P
CHORD ORGAN 9. Using black buttons, find highest chord			P
10. Using black buttons, find lowest chord			P
11. Play some short notes, then gradually get longer			D
12. Play long notes, then gradually get shorter			D
SIX RESONATOR BELLS 13. Find highest pitch			P
14. Arrange in order and play from low to high			P
15. Play fast from low to high*			D
16. Play slowly from high to low*			D
17. Play loudly from low to high*			L
18. Play softly from high to low*			L

*Do not consider correctness of pitch on these items.

Manipulative Measure, Form 3

Name _____

TESTING	ITEM	RIGHT	WRONG
D	DRUM 1. Begin playing fast and gradually get slower.		
D	2. Begin playing slowly and gradually get faster.		
L	3. Play softly, loudly, medium.		
D	CHORD ORGAN 4. Begin playing short notes and play each one a little longer.		
D	5. Begin playing long notes, and as you play have each note get a little shorter.		
P	6. Using black buttons - find the lowest sounding chord.		
P	7. Using black buttons - find the highest sounding chord.		
P	RESONATOR BARS (4) 8. Find highest bar.		
P	9. Arrange the bars in order and play from lowest to highest.		
L	10. Start at one end softly and make each one louder.		
L	11. Three mallets on wood block: Which one is loudest?		
P	12. Ukulele: Play strings in order from lowest to highest.		
L	13. Tape recorder: Find the softest position.		
D	14. Metal Square, finger cymbal, resonator bell: Which one sounds for the shortest time?		
L	15. 3 Jingle Bells: Which set is loudest?		
P	16. 4 Metal Plates: Play from highest to lowest?		
L	17. 3 Maracas: Which is softest?		
D	18. Drum, triangle, resonator bells: Which sound is medium in length?		

Manipulative Measure, Form 4

Name _____

TESTING		ITEM	RIGHT	WRONG
D	1.	On drum: Begin playing fast and gradually get slower.		
D	2.	On drum: Begin playing slowly and gradually get faster.		
P	3.	On chord organ: Using black buttons, find the lowest sounding chord.		
P	4.	Same: Find the highest sounding chord.		
P	5.	Resonator bars (4): Find the highest sound.		
P	6.	Resonator bars: Arrange in order and play from the lowest to the highest.		
L	7.	Three mallets on woodblock: Play from loudest to softest.		
P	8.	Ukulele: Play strings in order from lowest sounding to highest sounding.		
L	9.	Tape recorder: Find the soft, medium, and loudest positions.		
D	10.	Metal square, finger cymbal, resonator bell (F): Which one sounds for the shortest time?		
L	11.	Three jingle bells: Play in order from softest to loudest.		
P	12.	Four metal plates: Play from highest to lowest (child may rearrange these if it helps).		
L	13.	Three maracas: Play in order from loudest to softest.		
D	14.	Drum, triangle, resonator bells: Which sound is medium in length?		
L	15.	Metal square, finger cymbal, triangle: Play in order from soft to loud.		
D	16.	Copper pipes (7, 6, 2): Which pipe sounds for the longest time?		
D	17.	Metal plates (three, excluding the loudest): Which have the longest and shortest sounds?		
L	18.	Metal square, woodblock, horseshoe, triangle, and resonator bell: Choose the three loudest and play in order of loudness from loud to soft.		

Musical Sources of Items and Dimension Measured
Overt Measure, Form 1

	Musical Source	Dimension Measured
1.	Handel: Royal Fireworks Music, La Rejoissance	loudness
2.	Handel: Water Music, Passepied	loudness
3.	Ponchielli: La Gioconda, Dance of the Hours	duration
4.	Herold: Zampa, Overture	loudness
5.	Herold: Zampa, Overture	duration and loudness
6.	Tchaikowsky: Marche Slav	loudness
7.	Von Suppe: Poet and Peasant, Overture	duration
8.	Saint-Saens: Carnival of the Animals, Fossils	loudness
9.	Saint-Saens: Carnival of the Animals, Elephants	duration
10.	Stravinsky: Petrouchka, Russian Dance	duration
11.	Prokofiev: Lieutenant Kije, Kije's Wedding	pitch
12.	Kabalevsky: The Comedians, March	loudness
13.	Villa-Lobos: The Little Train of the Caipira	duration
14.	Khachaturian: Masquerade Suite, Galop	duration and loudness
15.	Sullivan: Iolanthe, Overture	loudness
16.	Sullivan: Pinafore, Overture	pitch
17.	Sullivan: Pinafore, Overture	duration

Answer Sheet, Overt Measure, Form 1

Name ____________________
(Last Name First)

Item No.	Coded Response (Check any observed Response): Large to Small Movements	Small to Large	High to Low	Low to High	Fast to Slow	Slow to Fast	Smooth to Jerky	Jerky to Smooth	Even to Uneven	Uneven to Even	UNCODED RESPONSE OR GENERAL COMMENT

Musical Sources of Items and Dimension Measured
Overt Measure, Form 2

	Musical Source	Dimension Measured
1.	Tchaikowsky: Marche Slav	loudness
2.	Von Suppe: Poet and Peasant, Overture	duration
3.	Saint-Saens: Carnival of the Animals, Elephants	duration
4.	Stravinsky: Petrouchka, Russian Dance	duration
5.	Prokfiev: Lieutenant Kije, Kije's Wedding	pitch
6.	Kabalevsky: The Comedians, March	loudness
7.	Villa-Lobos: The Little Train of the Caipira	duration
8.	Khachaturian: Masquerade Suite, Galop	duration and loudness
9.	Sullivan: Iolanthe, Overture	loudness
10.	Sullivan: Pinafore, Overture	pitch
11.	Sullivan: Pinafore, Overture	duration

Answer Sheet, Overt Measure, Form 2

	1	2	3	4	5	6	7	8	9	10	11
No response											
Larger											
Smaller											
Higher											
Lower											
Faster											
Slower											
Smoother											
Less Smooth											
Description of Uncoded changes or General remarks											
Is change made at proper time? Yes											
No											
Is change appropriate to music? Yes											
No											
Changes in music	softer	faster	stac-cato to legato	slower	smoother lower			slower faster louder		high low high	slower
	1	2	3	4	5	6	7	8	9	10	11

Musical Sources of Items and Dimension Measured
Overt Measure, Form 3

	Musical Source	Dimension Measured
1.	Tchaikowsky: Marche Slav	loudness
2.	Von Suppe: Poet and Peasant, Overture	duration
3.	Stravinsky: Petrouchka, Russian Dance	duration
4.	Beethoven: Symphony No. 3, Third Movement	loudness
5.	Sullivan: Pinafore, Overture	duration
6.	Tchaikowsky: Overture 1812	pitch
7.	Moussorgsky-Ravel: Pictures at an Exhibition	pitch
8.	Mozart: The Marriage of Figaro, Overture	loudness
9.	Wagner: Die Meistersinger, Prelude to Act I	pitch

Answer Sheet, Overt Measure, Form 3

Item	No response	Change as music changes		Change appropriate to music		Larger	Smaller	Higher	Lower	Faster	Slower	Smoother	Less smooth	Louder	Softer	Remarks or uncoded responses
		Yes	No	Yes	No											
1 softer																
2 faster																
3 slower																
4 louder																
5 slower																
6 higher																
7 lower																
8 louder																
9 higher																

Musical Sources of Items and Correct Responses
Overt Measure, Form 4

	Musical Source	Correct Response
1.	Tchaikowsky: March Slav	softer
2.	Von Suppe: Poet and Peasant, Overture	faster
3.	Stravinsky: Petrouchka, Russian Dance	slower
4.	Beethoven: Symphony No. 3, Third Movement	louder
5.	Sullivan: Pinafore, Overture	slower
6.	Glinka: Russlan and Ludmilla, Overture	higher
7.	Liszt: Hungarian Rhapsody No. 2	lower
8.	Mozart: The Marriage of Figaro, Overture	louder
9.	Bartok: Concerto for Orchestra, Second Movement	higher

(For Answer Sheet see Appendix C, p.163.)

Appendix C: Measures Used in Main Study

Verbal Measure

Name _____

School _____

Date _____

Choose the best answer to complete each statement. Fill in the square beside the answer you think is best. Choose only one answer for each number.

Sample

If you sing out of tune, this means

- ☐ you sing too fast.
- ☐ you sing too softly.
- ☐ you sing too loudly.
- ☐ you sing higher or lower than the right notes.
- ☐ you sing faster or slower than the class.

---

1. As a band in a parade gets nearer to you,

   - ☐ the music seems louder.
   - ☐ the music seems higher.
   - ☐ the music seems louder and faster.
   - ☐ the music seems higher and faster.
   - ☐ the music seems lower and slower.

2. If you arranged five tones in order from highest to lowest, the lowest tone would be

   - ☐ louder than the highest.
   - ☐ shorter than the highest.
   - ☐ different from the highest.
   - ☐ the same as the highest.
   - ☐ softer than the highest.

3. If you sang a song slower and higher the second time,

- ☐ it would be softer.
- ☐ it would be louder.
- ☐ it would take less time.
- ☐ it would take more time.
- ☐ it would take the same amount of time.

4. A song for marching would be

- ☐ faster than a lullaby.
- ☐ slower than a lullaby.
- ☐ soft and slow.
- ☐ loud and slow.
- ☐ high and slow.

5. A short note could be

- ☐ loud.
- ☐ soft.
- ☐ high.
- ☐ fast.
- ☐ any of these.

6. When music gets softer we hear

- ☐ higher sound.
- ☐ lower sound.
- ☐ slower sound.
- ☐ less sound.
- ☐ longer sound.

7. If you played a song higher and softer the second time,

- ☐ it would sound bigger.
- ☐ it would sound smaller.
- ☐ it would take less time.
- ☐ it would take more time.
- ☐ it would use more notes.

8. If you sang a song faster and louder the second time,

- ☐ it would take less time.
- ☐ it would take more time.
- ☐ it would be higher.
- ☐ it would be lower.
- ☐ any of these.

9. Pick the best body movement to show someone that music gets lower.

- ☐ clapping
- ☐ stooping
- ☐ stop moving
- ☐ stretching
- ☐ stepping

10. In marching we step to the

- ☐ loudness of the music.
- ☐ tune of the music.
- ☐ lowness of the music.
- ☐ beat of the music.
- ☐ any of these.

11. When music gets louder,

- ☐ we hear more sound.
- ☐ we hear less sound.
- ☐ we hear higher sound.
- ☐ we hear lower sound.
- ☐ we hear faster sound.

12. A loud note could be

- ☐ high.
- ☐ low.
- ☐ long.
- ☐ short.
- ☐ any of these.

13. As a melody goes up, the tones must become

- ☐ softer.
- ☐ louder.
- ☐ higher.
- ☐ lower.
- ☐ faster.

14. If you played a song louder and lower the second time,

- ☐ you would finish it sooner.
- ☐ there would be less sound.
- ☐ there would be more sound.
- ☐ it would be faster.
- ☐ it would be slower.

15. High music is most like

- ☐ thunder.
- ☐ singing birds.
- ☐ a cow's moo.
- ☐ a rumbling truck.
- ☐ roaring cannons.

16. Slow music is most like

- ☐ running.
- ☐ skipping.
- ☐ creeping.
- ☐ galloping.
- ☐ spinning.

17. Which of these instruments can sound several notes <u>at the same time</u>?

- ☐ drum
- ☐ piano
- ☐ trumpet
- ☐ flute
- ☐ voice

18. A high note <u>could</u> be

- ☐ loud.
- ☐ soft.
- ☐ long.
- ☐ short.
- ☐ any of these.

arranged

movement

instruments

melody

lullaby

stretching

music

Musical Sources of Items and Correct Responses
Listening Measure

	Musical Source	Correct Response
	Sample A Sullivan: *Iolanthe*, Overture	louder
	Sample B Beethoven: *Leonore Overture No. 2*	lower
1.	Sullivan: *Pinafore*, Overture	softer
2.	Prokofiev: *Classical Symphony*, Movement 1	higher
3.	Bach: *Toccata in D Minor*	lower
4.	Beethoven: *Symphony No. 3*, Movement 3	louder
5.	Handel: *Water Music*, Hornpipe	slower
6.	Lecuona: *Malaguena*	faster
	Sample C Beethoven: *Egmont Overture*	higher and louder
7.	Stravinsky: *Petrouchka*, Russian Dance	softer and slower
8.	Beethoven: *Leonore Overture No. 2*	higher and louder
9.	Tchaikowsky: *Swan Lake*, Czardas	softer and faster
10.	Prokofiev: *Lieutenant Kije*, Kije's Wedding	lower and slower
11.	Beethoven: *Consecration of the House*	higher and faster
12.	Lizst: *Grande Galop Chromatique*	lower and louder
	Sample D Holst: *The Planets*, Jupiter	lower
	Sample E Brahms: *Variations on a Theme by Haydn*	softer
13.	Brahms: *Variations on a Theme by Haydn*	faster
14.	Tchaikowsky: *Swan Lake*, Little Swans	higher
15.	Tchaikowsky: *Swan Lake*, Waltz	louder
16.	Gounod: *Faust*, Dance of the Nubians	lower
17.	Ravel: *Ma Mere L'Oye*, Pavanne	softer
18.	Brahms: *Hungarian Dance No. 6*	slower

Answer Sheet, Listening Measure

Name _____

Directions:

We will hear some short musical examples. Each one will change in some way. Listen to the complete example and decide on the most important change. Read through the possible answers on your paper, but before marking your choice, listen again as the same example is repeated. After hearing the example the second time, fill in the square beside the answer you think is right. Be sure you fill in only one square for each example.

SAMPLE A

- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer
- ☐ higher

SAMPLE B

- ☐ louder
- ☐ higher
- ☐ lower
- ☐ faster
- ☐ slower

---

1. ☐ faster
   ☐ slower
   ☐ louder
   ☐ softer
   ☐ lower

2. ☐ faster
   ☐ slower
   ☐ louder
   ☐ higher
   ☐ lower

3. ☐ louder
   ☐ softer
   ☐ lower
   ☐ higher
   ☐ faster

4. ☐ louder
   ☐ softer
   ☐ faster
   ☐ slower
   ☐ lower

5. ☐ softer
   ☐ higher
   ☐ lower
   ☐ faster
   ☐ slower

6. ☐ faster
   ☐ slower
   ☐ softer
   ☐ higher
   ☐ lower

In this section each answer has two parts. For example, the music may get higher and slower, or faster and louder. Choose the answer that has both parts right, but mark only one square for each example.

SAMPLE C

- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

7.
- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

8.
- ☐ higher and softer
- ☐ higher and louder
- ☐ higher and slower
- ☐ lower and louder
- ☐ lower and faster

9.
- ☐ slower and softer
- ☐ slower and louder
- ☐ faster and softer
- ☐ faster and louder
- ☐ higher and louder

10.
- ☐ higher and softer
- ☐ higher and louder
- ☐ higher and slower
- ☐ lower and slower
- ☐ lower and faster

11.
- ☐ lower and louder
- ☐ lower and faster
- ☐ higher and faster
- ☐ higher and slower
- ☐ louder and slower

12.
- ☐ lower and softer
- ☐ lower and louder
- ☐ higher and softer
- ☐ slower and softer
- ☐ higher and louder

In this group each example has two different parts. Listen to both parts and decide how the second part is most different from the first. Mark the square beside the word which best describes this difference. Mark only one square for each example. Listen carefully as these are played, because they will not be repeated.

SAMPLE D

- ☐ faster
- ☐ slower
- ☐ higher
- ☐ lower
- ☐ louder

SAMPLE E

- ☐ louder
- ☐ softer
- ☐ higher
- ☐ slower
- ☐ faster

13.
- ☐ lower
- ☐ louder
- ☐ softer
- ☐ faster
- ☐ slower

14.
- ☐ lower
- ☐ higher
- ☐ slower
- ☐ faster
- ☐ softer

15.
- ☐ louder
- ☐ slower
- ☐ faster
- ☐ higher
- ☐ lower

16.
- ☐ lower
- ☐ higher
- ☐ faster
- ☐ slower
- ☐ softer

17.
- ☐ lower
- ☐ higher
- ☐ faster
- ☐ louder
- ☐ softer

18.
- ☐ lower
- ☐ faster
- ☐ slower
- ☐ louder
- ☐ softer

Directions for Administration, Listening Measure

DIRECTIONS	ON TAPE
Distribute answer sheets. Make certain each child has a pencil with an eraser. Instruct children to write their names in the space provided on the first page. Start tape.	"At the top of your answer sheet there are some directions. Read them silently to yourself while I read them to you: 'We will hear some short musical examples. Each one will change in some way. Listen to the complete example and decide on the most important change. Read through the possible answers on your paper, but before marking your choice, listen again as the same example is repeated. After hearing the example the second time, fill in the square beside the answer you think is right. Be sure you fill in only one square for each example.' Let's try a sample. Listen to Sample A and hear how it changes." (Sample A) "Look at your answer sheet and decide which answer best describes how the music changed." (Pause) "Listen again before marking your answer to see whether you were correct." (Sample A repeated) "Mark you answer for Sample A."
Stop tape. Ask: "WHICH WORD DESCRIBES THE MOST IMPORTANT CHANGE?"	

DIRECTIONS	ON TAPE
Allow children to answer (correct answer is 'louder'). Say: "FILL IN THE SQUARE BESIDE 'LOUDER'. IF YOU HAD A DIFFERENT ANSWER, ERASE IT." After all have corrected Sample A, Say: "LET'S TRY ANOTHER SAMPLE." Start tape.	"Listen to Sample B." (Sample B) "Decide what you think is the most important change. Look for your answer on the sheet, but listen again before marking your answer." (Sample B repeated) "Mark your answer for Sample B."
Stop tape. After children have marked answers, Ask: "WHAT DID YOU MARK?" After children have answered, Say: "LOWER IS RIGHT. ARE THERE ANY QUESTIONS ABOUT HOW TO DO THIS?" Answer any questions about procedure, then, Say: "LET'S BEGIN WITH NUMBER ONE." Start tape.	(Items 1 through 6) "Turn to the directions at the top of page 2." (Pause) "In this section each answer has two parts. For example, the music may get higher and slower, or faster and louder. Choose the answer that has both parts right, but mark only one square for each example." "Listen to a sample. Sample C." (Sample C)

DIRECTIONS	ON TAPE
	"Decide the two most important ways the music changed and find the answer; but listen again before marking the answer."
	(Sample C repeated)
	"Mark your answer for Sample C."
Stop tape. After children have marked answers, Ask: "WHICH ANSWER DID YOU MARK?" Allow children to answer (higher and louder) Say: "BE SURE YOU HAVE FILLED IN THE ONE SQUARE BESIDE THE TWO WORDS 'HIGHER AND LOUDER.' MARK ONLY THE ONE SQUARE WITH BOTH PARTS CORRECT. ARE THERE ANY QUESTIONS?" After answering any questions, Start tape.	
	(Items 7 through 12)
	"Turn to the directions at the top of page 3."
	(Pause)
	"In this group each example has two different parts. Listen to both parts and decide how the second part is most different from the first. Mark the square beside the word which best describes this difference. Mark only one square for each example. Listen carefully as these are played, because they will not be repeated. Let's try a sample. Sample D."
	(Sample D)
	"Mark the square beside the word which tells how the second part was different from the first part."
Stop tape. Ask: "WHICH IS THE CORRECT ANSWER?" Allow children to answer (lower).	

DIRECTIONS	ON TAPE
After children have corrected answers, Say: "LET'S TRY ANOTHER SAMPLE." Start tape.	
	"Sample E." (Sample E) "Mark your answer for Sample E."
Stop tape. Give children time to mark an answer. Ask: "WHICH ANSWER DID YOU MARK?" Allow children to answer. Say: "SOFTER IS RIGHT." Allow children to correct answer. Ask: "ARE THERE ANY QUESTIONS?" After answering questions, Say: "BE SURE TO TELL HOW THE WHOLE SECOND PART IS DIFFERENT FROM THE WHOLE FIRST PART, NOT JUST HOW THE BEGINNING OF THE SECOND PART IS DIFFERENT FROM THE END OF THE FIRST PART. LET'S GO AHEAD WITH NUMBER 13. REMEMBER, THERE ARE TWO DIFFERENT PARTS FOR EACH EXAMPLE, AND THESE ARE PLAYED ONLY ONCE." Start tape.	
	(Items 13 through 18)
Stop tape. Allow time for children to answer. Say: "FOLD THE PAGE BACK SO THAT THE SHEET WITH YOUR NAME IS ON TOP." Collect the papers.	

Answer Sheet, Manipulative Measure

Name _____

Measuring	Item	Right	Wrong
D	1. Drum: Begin playing fast and gradually get slower.		
D	2. Drum: Begin playing slowly and gradually get faster.		
P	3. Chord Organ: Using black buttons, find the lowest sounding chord.		
P	4. Same: Find the highest sounding chord.		
L	5. Metal square, resonator bell, triangle, (all with rubber mallet): Play in order from loudest to softest.		
L	6. Metal square, resonator bell, horseshoe (all with rubber mallet): Play in order from softest to loudest.		
P	7. Resonator bars: Play the highest sounding.		
P	8. Resonator bars: Play in order.		
D	9. Drum, triangle, resonator bell: Play the one which is medium in length.		
D	10. Metal square, finger cymbal, resonator bell: Play the one which sounds for the shortest time.		
L	11. Metal square, finger cymbal, resonator bell (with rubber mallet): Play in order from softest to loudest.		
L	12. Chord organ, metal square, resonator bell (wooden mallet): Play in order from loudest to softest.		
P	13. Ukulele: Play strings in order from lowest to highest.		
P	14. (4) Metal Plates: Play from highest to lowest.		
D	15. (3) Metal Plates: Play the one which has the shortest sound.		
D	16. (3) Copper pipes: Play the pipe which sounds for the longest time.		
L	17. Three notes on chord organ: Play in order from loudest to softest.		
L	18. Tape recorder: Show the softest, medium, and loudest positions.		

Directions for Administration, Manipulative Measure

The test administrator greets the child. The exact manner of greeting depends upon the appearance and actions of the child and cannot be specified. The administrator's manner should be friendly, aiming to relax the child and put him at ease.

The test administrator will then say: I AM GOING TO ASK YOU TO PLAY THESE INSTRUMENTS. YOU MAY TRY THEM AS MANY TIMES AS YOU WISH. THEN, WHEN YOU HAVE DECIDED UPON YOUR ANSWER, SAY, "THIS IS MY ANSWER," AND PLAY YOUR ANSWER FOR ME. DO YOU HAVE ANY QUESTIONS?

SUGGESTIONS:

1. The order of placement of the instruments should be constant. The instruments should be numbered so that they can be arranged in this order before each child enters.

2. The test administrator should avoid comments that might cue the child. When a child looks for approval, the test administrator should say, "Remember, when you are ready, say, 'This is my answer,' and then play your answer." Administrator should avoid saying, after child has indicated his choice, "Is that your answer?" or "Are you sure?" He should also avoid reinforcing any answers.

3. If test administrator has not heard the answer, or is not sure he has remembered it correctly, he should say, "Will you repeat that, please? I didn't hear it."

4. Test administrator on the first few questions should remind pupil to say "This is my answer," and then play his answer. He may, if necessary, also remind the pupil that he may try the instruments as much as he wants before he arrives at an answer.

Classroom Orientation, Overt Measure

After verbal and listening tests are complete, before leaving:

ASK: Have you ever thought of the ways you could move to music?

We are going to play some music. Think of some ways you could move.

PLAY EXCERPT - "Vienna Maidens" by Ziehrer, on Listening tape, about 10 spaces after the end of the Listening items.

After playing, ASK:

How could you move to that music?

Get as many responses as possible from class--use raised hands for calling on students to answer. (Possible responses--swinging, walking, turning, dancing, skipping, etc.)

ASK: What other things could you do to different music?

After responses are heard, SAY:

Listen to this music again and decide whether the biggest change was: higher, lower, softer, faster, or slower, louder.

PLAY EXCERPT AGAIN

ASK: What was the biggest, most important change? (Answer should be higher)

SAY: Later we will be asking some of you boys and girls to show us the ways you would move to different kinds of music. Remember some of the ways we talked about this morning.

Thank you very much for helping us with this project. We'll see some of you later.

Musical Sources of Items and Correct Responses
Overt Measure

	Musical Source	Correct Response
1.	Tchaikowsky: Marche Slav	softer
2.	Von Suppe: Poet and Peasant, Overture	faster
3.	Stravinsky: Petrouchka, Russian Dance	slower
4.	Beethoven: Symphony No. 3, Third Movement	louder
5.	Waldteufel: Intermezzo	slower
6.	Lehar: Gold and Silver Waltz	higher
7.	Liszt: Hungarian Rhapsody No. 2	lower
8.	Mozart: Marriage of Figaro, Overture	louder
9.	Bartok: Concerto for Orchestra, Second Movement	higher

Answer Sheet, Overt Measure

Name _____

	No Response	No Change	Change when music Changes	Correct verbal Response	Incorrect verbal Response		Remarks
1.						softer	
2.						faster	
3.						slower	
4.						louder	
5.						slower	
6.						higher	
7.						lower	
8.						louder	
9.						higher	

Directions for Administration, Overt Measure

"In your class we talked about the ways people move to music. Now we are going to play some music for you. Listen the first time and think how you could move to the music. It may make you think of running, skipping, walking, jumping, hopping, stretching, bending, stooping, reaching up, turning, or something else. You can use your hands, feet, your head, or your whole body.

"The second time the music plays, do what the music makes you think of doing. When the music changes, try to show us by your movement how the music changes."

PLAY EXAMPLE THE FIRST TWO TIMES, INDICATING TO THE CHILD (IF NECESSARY) THAT HE SHOULD MOVE THE SECOND TIME. ENCOURAGE HIM TO MOVE IF HE SEEMS HESITANT. AFTER PLAYING THE SECOND TIME, STOP THE TAPE (ON NUMBER ONE ONLY) TO SAY:

"Listen again and then tell the biggest change. Tell whether the biggest change is higher, lower, faster, slower, louder, or softer. You can look at this card to choose your answer."

INDICATE CARD. START TAPE AGAIN. IF THE CHILD IS SLOW IN ANSWERING, IT MAY BE NECESSARY TO STOP THE TAPE BEFORE NUMBER TWO BEGINS. IF TAPE IS STOPPED, SAY:

"There will be nine of these altogether. Listen, move, listen again, and answer. Go ahead with Number 2 now."

START TAPE, STOPPING ONLY IF REQUIRED FOR THE ORAL ANSWER.

Stimulus Card, Overt Measure

higher

lower

louder

softer

faster

slower

Appendix D: Information Forms Sent to Schools in Main Study Sample

The Pennsylvania State University
Department of Music Education

GENERAL INFORMATION

Time, Facilities, Personnel, and Related Information

Description of Measuring Instrument

The measuring instrument is divided into four submeasures: Listening, Verbal, Manipulative, and Overt. The first two submeasures are administered in combined form as a group measure in one classroom session; the remaining two submeasures are administered separately to individual pupils.

Group Measure

Time Required: Approximately one hour. Testing will be done within school hours. We need information on opening dates of school in the fall, any dates when testing will be inconvenient, and daily school hours.

Space Required: Classroom situation, children seated at regular desks, separated if possible.

Equipment Required: Pencil with eraser for each child. For the listening measure an electrical outlet is needed for the tape recorder. A quiet environment will contribute to the validity of this measure.

Individual Measures

Time Required: Total time 30 minutes per student, with two students out of class at any one time.

Space Required: Two rooms, one room large enough for the child to move around actively and freely, the other room smaller, perhaps office size.

Equipment & Facilities Required: Both rooms must have electrical outlets and a table or desk cleared for use. Other necessary equipment will be furnished by the research team.

Order of Administering Measures

Group measure first; individual measures second.

Test Scores and Other Information

Data Needed: Most recent IQ and standardized reading scores, names of tests used, and approximate dates administered.

Class roll of classes selected for measurement.

GENERAL INFORMATION (continued)

Names of children with other than normal hearing who are involved in measurement sample.

Selection of Students

The research team will make a random selection from all available classes and inform you which class or classes from your district will be included in the sample.

- - - - - - - - - - - - - - -

For your information, this measure has been administered to approximately sixty children in the pilot and pre-pilot programs. Parts of the measure have been given to approximately fifty additional children. There are indications that the testing program has been enjoyable for the participants. The members of the research team, all experienced in teaching children, make every effort to provide an atmosphere in which the children feel comfortable.

OE MUSICAL CONCEPTS
Project 2934 OE-6-10-002
The Pennsylvania State University
Department of Music Education

Research Team: Dr. Frances M. Andrews,
Dr. Ned C. Deihl, Grace E. Laverty,
Cathy J. Cobes, A. Peter Costanza

QUESTIONNAIRE: Part I*

1. Official name of school system _____

2. Name of chief administrator _____

   Official title _____

   Address _____

3. Name of Music Supervisor** _____

   Summer address _____

4. Opening date of school _____

5. Dates not desirable for visit _____

6. School day hours _____ A.M. to _____ P.M.

7. We will need IQ scores and standardized reading test scores. Do you prefer that we extract these scores when we visit the schools, or will it be more convenient for you to provide them? _____

8. Please list by room number, by teacher's name, or by some other means all available 4th grade classes in buildings with adequate space for the measurement procedure. It is important that homogeneous and heterogeneous groupings be listed in the appropriate columns. Please exclude Special Education classes. SEE EXAMPLE BELOW.

EXAMPLE

Building	Location	Heterogeneous Classes	Homogeneous Classes: How Grouped?	Approx. No. of Pupils in Class	Comments
Lincoln	Washington Twp.		Room 211: Miss Hale IQ's above 140	30	Alert, interested; good situation
Lincoln	Washington Twp.	Room 212		25	
Franklin	Adams Borough	Room 7		40	Uncooperative atmosphere
Stevens	Polk City	Miss Smith Room 3		20	Pupils are predominantly rural
Stevens	Polk City	Mrs. Warner Room 4		25	Slum area; Title I school
Stevens	Polk City		Room 213: Low achievement scores in reading	30	

*To be completed by Music Supervisor or other designated person.
**If a person other than Music Supervisor is school coordinator for this study, insert his name and address instead of Music Supervisor.

QUESTIONNAIRE: Part I (continued)

Form for Describing Available Classes

Building	Location	Heterogeneous Classes	Homogeneous Classes: How Grouped?	Approx. No. of Pupils in Class	Comments

OE MUSICAL CONCEPTS
Project 2934 OE-6-10-002

The Pennsylvania State University
Department of Music Education

Part II

The following information should be supplied by the chief school administrator or his administrative assistant.

Please give this sheet to the proper person, asking that he complete and return the form to us.

Name of school system _____

Location _____

School enrollment:

Secondary (7 through 12) _____

Elementary (1 through 6) _____

Total 4th grade _____

Predominant population characteristics of the school district (in percentages):

Urban _____ Surburban _____ Rural _____

Approximate per cent of parents in the following general categories:

Professional - Business _____ Semi-Professional (office workers) _____

Skilled Labor (factory workers) _____ Unskilled Labor _____

Agriculture _____

Assessed property value of the school district _____

Political subdivisions of the school district, listed by name:

Cities _____

Towns _____

Boroughs _____

Townships _____

Please list any Title I schools in the school district.

Return to:
Mrs. Diane Gold
Music Education Department
263 Chambers Building
University Park, Pennsylvania

Completed by: _____
Official title _____
Address _____
Date _____

Appendix E: Computer Programs Used in Statistical Analyses

ITANL

ITEM ANALYSIS

11.0.014

Walter Dick
Richard E. Spencer

The Pennsylvania State Univ.

Office of Examination Services

May 25, 1964

General Description:

1. This program computes for each test item: the biserial correlation*, the point-biserial and the t-value associated with it, the difficulty level (p), the mean total score of the students getting the item correct, and the mean total score of the students who miss the item.

2. Summary information including the mean difficulty index, mean biserial correlation, an estimate of interitem correlation, the standard error of the correlation, Kuder-Richardson #20 reliability, and test mean, variance and standard deviation are also computed.

Output:

1. For each item: item number, correct answer, a table indicating the frequency with which each of the answer choices was selected by each subgroup in the test population, a total frequency with which each answer choice was selected; each of the statistics described in section 1 of General Description and, if desired, an evaluation of item difficulty, item discrimination, and a total item evaluation.

2. For the total test: tables of frequency distributions of the test items with respect to item difficulty and item discrimination, plus the statistics described in section 2 of the General Description.

Definitions:

In this description, the following definitions apply:

Data deck = a group of card representing the responses of all the testees on 50 or fewer items.

Run = the processing of all the data decks which have been submitted. This would include up to 17 data decks if the cards are hand punched. There is only one data deck per run if Digitek Cards are used.

Capabilities and Limitations:

1. Only one test may be analyzed per run.

2. Any number of students may be used.

*Computational procedure from Computation Center program S11.3.006, BISER. J. E. Singer

3. Total score or criterion score of any individual cannot exceed 999.

4. If the tests have been scored by the Digitek Optical Scanner, the test cannot include more than 150 items.

5. If the test data cards are hand punched, any number of items up to 900 may be processed on one run (see section on Data Preparation).

6. The population which has been tested may be subdivided into two, three, four or five groups in order to check the responses selected by various subgroups. For example, the testees could be subdivided into males and females, or in terms of their total score, i.e. lowest one-fifth, second fifth, middle fifth, fourth fifth and highest fifth. This subgrouping does not effect the item statistics and is not a procedural requirement of the program. (See section of Data Preparation below.) Subgrouping is not required.

7. No item can have more than one correct answer. If a test is being processed which does not necessarily have correct answers, arbitrary "correct answers" must be submitted.

8. There can be no more than 5 possible response alternatives to any test items. (If an "omit" is counted as a response alternative, there are then 6 possible responses.) See section 1 of Data Preparation below.

9. There can be no double punched card columns, except for Digitek prepared cards, and then only in card column 80.

Preparation of Data Cards:

1. Hand punched cards.

Hand punched data may be used in standard format, which is total score or criterion score in cols. 1 - 3 (right justified) and item responses (i.e. 1, 2, 3, 4, 5 or 0 for an omit) in cols. 4 - 53 with one response per column. Standard format is indicated with a "1" in col. 60 of the parameter card. Or, a format may be supplied by the user. This format must provide for reading the total score first and then up to 50 items, both of which are to be read in integer format, e.g. (7X, I3, 35I2/10X, 15I2). Format at object time is indicated with a "2" in col. 60 of the parameter card.

It is important to note that this program analyzes the data for up to 50 items at one time. If the data is to be hand punched, and, there are more than 50 items in the test, separate data decks of cards must be made for each set of 50 items. Although the data is punched separately for each set of 50 items in the test, the items will be sequentially numbered in the print-out and the summary data will be for the total number of test items.

The program also requires that five cards with "-1" punched in cols. 1 and 2 be used with each data deck of 50 items or less. (If format is being entered at object time, the (-1)'s must be punched in the same card cols. as the total score in punched.) These cards are used to indicate the end of data for various subgroups within the testee population. The fifth (-1) card indicates the end of the data for one data deck of items. If the data cards for 40 students are sorted on total score from lowest to highest, five subgroups could be formed by placing a (-1) card after the data cards for the 8th., 16th., 24th., 32nd., and 40th. students. If the population were to be dicotomized into males and females, a (-1) card would be inserted after the card for the last male student and 4 (-1) cards after the card for the last female. If no grouping is desired, 5 (-1) cards would be inserted after the data deck for up to 50 items. This process must be repeated for each data deck.

Approximately 18 data decks representing 50 items each may be submitted for one run when the cards are hand punched and only one print-out is wanted. If more copies are needed, the number of items to be processed must be reduced in order not to exceed maximum number of records.

2. Digitek punched cards.

If the item analysis cards have been punched by the Digitek Optical Scanner, the user is limited to analyzing 150 items per run. No matter how many items are being processed, there will always be 2 Digitek cards per student. If subgroups are to be used, these cards can be sorted appropriately with each student's two cards staying together.

Five cards with (-1) punched in cols. 1 and 2 must be inserted into the Digitek data deck. These cards are used to indicate the end of subgroups within the testee population. See paragraph 3 of hand preparation of data cards for examples. The identical procedure is used for the Digitek cards with the exception that there will only be one data deck regardless of the number of test items (up to 150). This is the exception to the definition of data deck which appears on page 1.

Input Deck:

If hand punched cards are used:

A. Parameter card
B. Correct answer card
C. Format card, if necessary
D. Data cards, including 5 (-1) cards
E. Repeat A-B-C-D as necessary
F. Blank card - end of run*

If Digitek cards are used:

A. Parameter card
B. Correct answer card
C. All data cards, including 5 (-1) cards
D. Repeat A and B as necessary
E. Blank card - end of run*

*Number of copies of the test summary which are needed should be punched in col. 65 of the end of run card.

Parameter Card:

Cols. 1-4	Test Code (Numeric) (Some Number must be punched).
Cols. 5-14	Course Name
Cols. 15-24	Instructor or identification
Cols. 30-35	Date (e.g. 121164 which will be printed 12/11/64)
Cols. 37-40	Total number of students (right justified)
Cols. 43-45	No. of items in this data deck - right justified. (This is not the total number of items in the run.)
Col. 50	Item evaluation option a. If wanted, punch a 1 b. If unnecessary, punch a 2
Col. 55	No. of answer choices to the test items (i.e. 2, 3, 4, or 5).
Col. 60	Punch 0 if Digitek cards are used. Punch 1 if data cards are hand punched in standard format, i.e. total score or criterion in cols. 1 - 3, item responses in cols. 4 - 53. Punch 2 if hand punched data cards are punched in other than standard format.
Col. 65	Number of print-out copies needed. If 1, punch 1, etc. Up to 9 copies can be printed.

Correct Answer Card:

Beginning with column 1, punch the correct answer for each item, one answer per column. Total number of items cannot exceed 50 for any one data deck. No identification is punched in this card. Each item can have only one correct answer.

Answer choices A, B, C, D, E, and omit = 1, 2, 3, 4, 5 and 0.

Format Card:

If there is a 2 column 60 of the parameter card, the user must supply his own format card for reading the data. The format must provide for reading up to (but no more than) 50 items, and must be in fixed-point form.
Example: (7X, 13, 3512/10X,1512)

Data Cards:

See the appropriate section under Data Preparation.

Maximum Time Estimate:

For up to 300 students, approximately 5 seconds per item; 300 to 600 students, 10 seconds per item, etc. For 200 students and 60 items, time - 60 x 5 = 300. If Digitek cards are being used, add 100 seconds.

Maximum Records Estimate:

Approximately 25 records per item per copy. With sixty items and 2 output copies, records - 25 x 60 x 2 = 3000.

WARNING - Maximum records for one computer run is 25,000.

Formulas Used:

1. $X = \frac{\Sigma X}{N}$

2. $\sigma^2 = \frac{N \Sigma X^2 - (\Sigma X)^2}{N (N - 1)}$

3. $\sigma \quad \sqrt{\sigma^2}$

4. $S_1 E_{cor.} = \sqrt{\frac{1}{N - 1}}$

5. $p = \frac{N \text{ correct}}{\sqrt{N}}$

6. $r_{bis} = \frac{\overline{X}_{cor} - \overline{X}_{incor}}{\sigma} * \frac{p.q}{y}$ * = multiply

7. $r_{p.bis} = r_{bis} * \frac{y}{\sqrt{pq}}$

8. $t_{p\text{-}bis} = \frac{r_{bis} * \sqrt{N-2}}{1-r^2_{p\text{-}bis}}$

9. Reliability (K-R #20) = $\left(\frac{K}{K-1}\right)\left(\frac{\sigma^2 - \Sigma pq}{\sigma^2}\right)$

10. Mean Interitem Correlation = $\left(\frac{\Sigma r_{bis}}{K}\right)^2$

11. $\bar{X}_{diff} = \frac{\Sigma p}{K}$

For interpretation of the output of this program, contact the Office of Examination Services.

RELIB

COMPLETE A.O.V. RELIABILITY
Internal Consistency

Walter Dick
Office of Examination Services

March, 1965
Pennsylvania State University

Description:

The program will compute the internal consistency reliability of test data. The coefficient of reliability is determined through the use of the analysis of variance. It is exactly equivalent to the Kuder-Richardson Formula 20 estimate (Guilford, 1954, p. 383). The data may be of two types: (1) binary coded (1 and 0) to indicate correct and incorrect responses, or (2) continuous (1, 2, 3, 4, 5, ...) as on an attitude questionnaire.

In addition, the program will compute the reliability of the data when item clusters are specified by the user. That is, if the items may be grouped in some meaningful way such as test items 1 - 5 refer to reading paragraph 1, items 6 - 10 to paragraph 2, etc., then a strata random and strata fixed reliability can be computed (Rabinowitz and Eikeland, 1964).

If a parallel test were constructed, and the same types of paragraphs were selected, then the strata fixed estimate of reliability would be selected. If, however, paragraphs from different content areas or with different characteristics had been selected, then the strata random estimate would have been selected.

The program also computes the internal consistency reliability of each of the strata which are identified by the user and indicates the product-moment correlation among the scores on each of the strata.

If there are strata within the data, the data must be submitted with the items in each strata punched consecutively.

If the over-all reliability of the data only is desired, an "item per strata" card, containing the total number of items, still must be inserted.

Limitations:

a. Up to 500 total items may be used
b. Up to 40 strata may be used
c. Number of students unlimited

Input Deck:

a. Object deck
b. Parameter card
c. Item per strata card
d. Format card
e. Data cards
f. Repeat b through e
g. End of run - blank card

Parameter Card:

Col.	1 - 4	No. of strata or item clusters (right justified)
Col.	5 - 8	Total no. of test items (right justified)
Col.	9 -12	No. of subjects (right justified)
Col.	16	= 1 for total reliability only = 0 for stratified reliability estimates also
Col.	21 -40	Problem identification

Item per Strata Card:

Indicate the number of items in each strata.

Col.	1-2	No. of items in strata 1 (right justified)
Col.	3-4	No. of items in strata 2 (right justified)
Col.	5-6	No. of items in strata 3 (right justified)
etc.		

Format Card: (required)

A format for reading the item responses in floating point mode must be submitted. If 50 items were submitted, punched in the first 50 cols., the format would be (50F1.0).

Data Cards:

Data cards must be punched to exactly conform to the format which is submitted.

Maximum Time:

One third second per card

Maximum Records Estimate:

50 for each test being analyzed.

References:

Guilford, J. P. Psychometric methods. New York: McGraw Hill, 1954.

Rabinowitz, W. and Eikeland, H. M. Establishing the reliability of tests with clustered items. Pedagogisk Forskning, 1964, p. 85-106.

THE PENNSYLVANIA STATE UNIVERSITY
Computation Center

COREL
11.3.003 (Revised)

SYMETRIC CORRELATION PROGRAM
FLOATING INPUT

M. E. Roberts
A. T. Wink

January 1964

GENERAL DESCRIPTION

This program computes correlation coefficients of all possible pairs of variables plus the means and standard deviations of all variables.

OUTPUT

Correlations are printed in a triangular matrix followed by tables of means and standard deviations. All values are rounded to 3 decimals.

If multiple regression or factor analysis is to be computed, sets of cards containing means, standard deviations and correlations are punched for variable 1 through the variable stated on the parameter card. Means and standard deviations are rounded to 3 decimals and correlations to 5.

CAPABILITIES AND LIMITATIONS

Number of variables may not exceed 105
Data, read with Standard Format, is 4-digit
Data size and/or decimal designation may be changed by inserting a format at object time
Decimals punched in data take precedence over format specification
Number of cases is unlimited, but if the number of cases is very large, accuracy will decrease as indicated under method.
If signs are punched they precede the data
Positive signs usually are not punched
Negative signs must be punched
Case numbers must be positive, non-zero, and generally distinct
Successive case numbers may never be identical
Cards per case are numbered sequentially starting with 1. Sequence within each case is checked.
If the total no. of variables can be punched on one card, case numbers and sequence numbers are not used. (See PARAMETER CARD - Col. 23)

METHOD

All computations are done in single precision floating point arithmetic. This provides 8-digit values for computing. Data of larger magnitude (4- or 5-digit data) or problems with large numbers of cases (over 1000) will develop output of lowered accuracy.

INPUT

A. PARAMETER CARD
B. FORMAT CARD (if format choice is "2")
C. DATA CARD
D. TRANSFER CARD (not used at end of final data set)
Repeat A, B, C, D as desired
E. END-OF-RUN CARD - Used only after end of F. D. S.

PARAMETER CARD

Right justify all values except problem name.
Unused columns may be unpunched or zeroes.

Col. 1-9 PARAMETER

Col. 12-14 Number of variables ($\leq$ 105)

Col. 16-18 Number of variables per card

Col. 21 Format choices are 1, 2, or 3.

1. Standard format applies: there are 2 possible forms depending upon whether Col.23 of parameter card is 0 or 1. (See Col. 23)
   a. If Col. 23 of parameter card is 0; Standard format is (2I4, 18F4.0)
      Data cards are then:
      Col. 1-4 Case number
      Col. 5-8 Card sequence number
      Col. 9-80 Eighteen 4-digit fields for variables
   b. If Col. 23 of parameter card is 1; Standard format is (20F4.0)
      Data cards are then:
      Col. 1-80 Twenty 4-digit fields for variables
2. Format card, inserted with this problem, applies.
3. Format card, most recently inserted, applies.
   Before choice 3 may be used the user must have inserted his own format card under choice 2 in the previous problem.

Col. 23 0 or 1

a. 0 if case and sequence numbers are necessary. Must be 0 is TNOV > No. of variables per card.
b. Must be 1 if total No. of variables = number of variables per card.

Col. 26-29 Number of last variable to be used for Factor Analysis or Multiple Regression. Means, standard deviations and correlations will be punched for all variables from 1 through number punched here. If no cards are desired, leave these columns blank. Must be 64 or less for MREG; 105 or less for FAN.

Col. 31-38 Problem name

FORMAT CARD

FORMAT (Ia, Ib, cFw.d)	if 0 in Col. 23 of parameter card

NOTE: Col. 1-6 must have the word FORMAT.

a = width of field for case numbers
b = width of field for card sequence numbers
c = number of variables per card. (parameter card Cols. 17-18)
w = width of field for each variable
d = number of places after decimal to be read.

FORMAT (cFw.d)	if 1 in Col. 23 of parameter card. c, w, d, are interpreted as above.

DATA CARDS

A B V_1 V_2	if 0 in Col. 23 of parameter card

A = case number - Must be positive
B = Card sequence number
V = Variables
All values right justified

V_1 V_2 V_3	if 1 in Col. 23 of parameter card

TRANSFER CARD

Col. 1-8 TRANSFER

END-OF-RUN CARD

Col. 1-3 END

EQUATIONS

$$\bar{X}_i = \frac{\Sigma X_i}{N}$$

$$\bar{X}_i = \sqrt{\frac{\Sigma X_i^2}{(N-1)} - \bar{X}_i^2 \quad \frac{N}{(N-1)}}$$

$$\sqrt{X_i X_j} = \frac{\Sigma X_i X_j - \bar{X}_i \Sigma X_j}{N \; X_i X_j}$$

DESCRIPTION OF CARD OUTPUT

If cards are requested, a single card deck is returned for each batch of problems regardless of the number of problems requiring cards.

This deck can be separated into problem decks by interpreting the first 60 columns (use standard panel) and reading Cols. 1-8 which contain problem name. The order of cards must not be changed.

Layout of Mean and Standard Deviation Cards

Cols.	1-8	Problem name
Cols.	9-17	Type Means Standard deviations
Cols.	18-20	Card sequence within type
Cols.	21-80	Six 10-digit fields for means and standard deviations

Layout of Correlation Cards

Cols.	1-8	Problem name
Cols.	9-13	Type - COR
Cols.	14-16	Card sequence within type
Cols.	17-80	Eight 8-digit fields for correlations

MAXIMUM TIME

300 seconds plus 20 seconds for each 100 data cards

MAXIMUM OUTPUT RECORDS

500

NOTE: The differences between this program and the original COR-F are that: (a) card sequence and case numbers are not necessary if the no. of variables = no. of variables per card. This option is explained under Col. 23 of the parameter card; (b) Standard deviations are computed using n-1 degrees of freedom rather than n degrees of freedom; and (c) control cards are mnemonic in designation.

OE 6000 (REV. 9-66)

DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION

ERIC REPORT RESUME

(TOP)

ERIC ACCESSION NO.	
CLEARINGHOUSE ACCESSION NUMBER	
RESUME DATE	08-10-67
P.A.	
T.A.	

IS DOCUMENT COPYRIGHTED? YES ☐ NO ☒

ERIC REPRODUCTION RELEASE? YES ☒ NO ☐

001

100 TITLE
101 Development of a Technique for Identifying Elementary School Children's
Musical Concepts.
102
103 Final Report Project No. 5-0233

200 PERSONAL AUTHOR(S)
Andrews, Frances M. Deihl, Ned C.

300 INSTITUTION (SOURCE)
Pennsylvania State University, University Park, Pa., College of Educ. SOURCE CODE

310 REPORT/SERIES NO.
OTHER SOURCE SOURCE CODE

320

330 OTHER REPORT NO.
OTHER SOURCE SOURCE CODE

340

350 OTHER REPORT NO.

400 PUBL. DATE 30-Sep-67 CONTRACT GRANT NUMBER OEC-6-10-002

PAGINATION ETC

500 192p.
501

RETRIEVAL TERMS

600 Musical concepts
601 Pitch
602 Duration
603 Loudness
604 Musical elements
605 Musical Concept Measures
606 Music Tests
Elementary music tests
Fourth-grade music

IDENTIFIERS

607 Battery of Musical Concept Measures

ABSTRACT

800 A multimodal technique (The Battery of Musical Concept Measures) was developed for
801 identifying elementary school children's concepts of pitch, duration, and loudness.
802 The Battery included four measures: Verbal, written responses to written stimuli;
803 Listening, written responses to multidimensional musical stimuli; Manipulative,
804 oral responses to subject-manipulated stimulus materials; Overt, overt movement
805 and oral responses to musical stimuli. Reliabilities were: Verbal, .71; Listen-
806 ing, .85; Manipulative, .66; Overt, .64. The group measures, Verbal and Listening,
807 were administered to 429 fourth-grade children from sixteen randomly-selected
808 classrooms in twelve representative Pennsylvania school districts; the individual
809 measures, Manipulative and Overt, were administered to 214 subjects randomly se-
810 lected from the total sample. Written scores for pitch, duration, and loudness
811 were derived from combining the group measures; nonwritten scores from combining
812 the individual measures. Reliabilities were: Written Pitch, .71; Written Dura-
813 tion, .69; Written Loudness, .60; Nonwritten Pitch, .65; Nonwritten Duration, .55;
814 Nonwritten Loudness, .34. Content validity was established by consensus of ex-
815 perts; construct validity was investigated in relation to a multitrait-multimethod
816 matrix. Correlations and partial correlations with IQ and reading scores and a
817 symmetric correlation matrix of derived scores are reported. A secondary analysis
818 of the Overt Measure showed no significant relationship between overt and oral
819 responses, and indicated some confusion by subjects in labeling musical changes.
820 Although further refinement is needed before the Battery is acceptable for practi-
821 cal use, the present measures are adequate for research purposes.
822