McDONALD, Dorothy Taylor, 1926-
THE IDENTIFICATION OF ELEMENTARY SCHOOL CHILDREN'S
MUSICAL CONCEPTS AS A FUNCTION OF ENVIRONMENT.

The Ohio State University, Ph.D., 1970
Education, general

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED
THE IDENTIFICATION OF ELEMENTARY SCHOOL CHILDREN'S
MUSICAL CONCEPTS AS A FUNCTION OF ENVIRONMENT

DISTRIBUTION
Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Dorothy Taylor McDonald, B.M.Ed., M.E.

# # # # # #

The Ohio State University
1970

Approved by

Henry L. Cady
Adviser
Department of Music Education
PLEASE NOTE:

Some pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS.
ACKNOWLEDGMENTS

The writer wishes to express her appreciation to the members of her committee, Dr. Henry L. Cady, Chairman, Dr. L. O. Andrews, Dr. Richard Hoppin, and Dr. H. Wayne Ramsey for their interest and assistance. The assistance of the chairman, Dr. Cady, is especially recognized.

Acknowledgment is due to all those persons who assisted in this study. The research would not have been possible without the cooperation of many children, their teachers, and their school administrators.

The assistance of Dr. Robert Ullman of the OSU Center for Evaluation and Measurement is acknowledged. The writer also wishes to express her appreciation to Dr. William Poland for making available the facilities of the School of Music Sound Laboratory for developing the audio-tape used in the study. Acknowledgment is due also to the Ohio State students and their instructors who participated in the study.
VITA

December 31, 1926 . . . Born - Randolph, Iowa

1949 . . . . . . . B.M. in Ed., The University of Nebraska, Lincoln, Nebraska


1963 . . . . . . . M.E., The University of North Carolina, Greensboro, North Carolina

1963-1964 . . . . Assistant Professor, Music Education, Greensboro College, Greensboro, North Carolina


1966-1968 . . . . Teaching Associate, Music Education, The Ohio State University, Columbus, Ohio

1968-1970 . . . . Instructor, Music Education, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Music Education

Music Education. Professor Henry L. Cady

Music History and Literature. Professor Richard L. Hoppin

Teacher Education. Professor L. O. Andrews

Elementary Education. Professor James Burr
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Problem</td>
<td></td>
</tr>
<tr>
<td>Problem Specified</td>
<td></td>
</tr>
<tr>
<td>Purposes of the Study</td>
<td></td>
</tr>
<tr>
<td>Sub-Purpose of the Study</td>
<td></td>
</tr>
<tr>
<td>Hypotheses</td>
<td></td>
</tr>
<tr>
<td>Assumptions</td>
<td></td>
</tr>
<tr>
<td>Delimitations</td>
<td></td>
</tr>
<tr>
<td>Definitions</td>
<td></td>
</tr>
<tr>
<td>II. REVIEW OF RELATED LITERATURE</td>
<td>15</td>
</tr>
<tr>
<td>Concepts and Concept Development</td>
<td></td>
</tr>
<tr>
<td>Concept-Learning and the Lower-Class Child</td>
<td></td>
</tr>
<tr>
<td>Musical Concepts and Listening Perception</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>III. THE DEVELOPMENT OF A TEST OF MUSICAL CONCEPTS</td>
<td>52</td>
</tr>
<tr>
<td>Procedures</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>IV. A PILOT STUDY OF TWO SCHOOLS</td>
<td>84</td>
</tr>
<tr>
<td>Procedures</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
</tr>
<tr>
<td>V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</td>
<td>106</td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>Conclusions and Recommendations</td>
<td></td>
</tr>
</tbody>
</table>

iv
<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>115</td>
</tr>
<tr>
<td>B</td>
<td>121</td>
</tr>
<tr>
<td>C</td>
<td>128</td>
</tr>
<tr>
<td>D</td>
<td>134</td>
</tr>
<tr>
<td>E</td>
<td>147</td>
</tr>
<tr>
<td>SELECTED BIBLIOGRAPHY</td>
<td>163</td>
</tr>
<tr>
<td>Table</td>
<td>Exploratory Trial; Form A; Range, Mean, Standard Deviation, and Reliability Estimate</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Exploratory Trial; Form A: Summary Item Difficulty</td>
</tr>
<tr>
<td>2</td>
<td>Exploratory Trial; Form A: Item Analysis</td>
</tr>
<tr>
<td>3</td>
<td>Exploratory Trial; Form A: Item Analysis</td>
</tr>
<tr>
<td>4</td>
<td>Trial I; Form A: Range, Mean, Standard Deviation, and Reliability Estimate</td>
</tr>
<tr>
<td>5</td>
<td>Trial I; Form A: Summary Item Difficulty</td>
</tr>
<tr>
<td>6</td>
<td>Trial I; Form A: Item Analysis</td>
</tr>
<tr>
<td>7</td>
<td>Trial II; Form B: Range, Mean, Standard Deviation, and Reliability Estimate</td>
</tr>
<tr>
<td>8</td>
<td>Trial II; Form B: Summary Item Difficulty</td>
</tr>
<tr>
<td>9</td>
<td>Trial II; Form B: Item Analysis</td>
</tr>
<tr>
<td>10</td>
<td>Trial II; Form B: Test Sections Mean Difficulty</td>
</tr>
<tr>
<td>11</td>
<td>Trial II; Form B: Dimension Mean Difficulty</td>
</tr>
<tr>
<td>12</td>
<td>Trial III; Form C: Range, Mean Standard Deviation, and Reliability Estimate</td>
</tr>
<tr>
<td>13</td>
<td>Trial III; Form C: Summary Item Difficulty</td>
</tr>
<tr>
<td>14</td>
<td>Trial III; Form C: Test Sections Mean Difficulty</td>
</tr>
<tr>
<td>15</td>
<td>Trial III; Form C: Dimension Mean Difficulty</td>
</tr>
<tr>
<td>16</td>
<td>Trial III; Form C: Item Analysis</td>
</tr>
<tr>
<td>17</td>
<td>Percentage of Correct Answers, Listening Test, Form C: Validation Trial</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>18. Listening Test; Form C: Means, Standard Deviations, Ranges, Reliability Estimates, and t Ratio, School A and School B</td>
<td>93</td>
</tr>
<tr>
<td>19. Listening Test; Form C: Mean Difficulty, School A and School B</td>
<td>94</td>
</tr>
<tr>
<td>20. Listening Test; Form C: Summary Item Difficulty Distribution, School A and School B</td>
<td>95</td>
</tr>
<tr>
<td>21. Listening Test; Form C: Test Sections Mean Difficulty, School A and School B</td>
<td>96</td>
</tr>
<tr>
<td>22. Listening Test; Form C: Dimension Mean Difficulty, School A and School B</td>
<td>96</td>
</tr>
<tr>
<td>23. Listening Test; Form C: Differences Between Proportions of Correct Answers in School A and School B</td>
<td>97</td>
</tr>
<tr>
<td>24. Means, Standard Deviations, t Values, Level of Significance; Other Variables, School A and School B</td>
<td>98</td>
</tr>
<tr>
<td>25. Relationship of Various Factors to Listening Test Scores in School A (Coefficients of Correlation)</td>
<td>99</td>
</tr>
<tr>
<td>26. Relationship of Other Factors to Listening Test Scores with Mental Age Scores Held Constant, School A (Partial Correlation Coefficients)</td>
<td>101</td>
</tr>
<tr>
<td>27. Summary: Relationship Between Listening Test Scores and Various Predictors: School A</td>
<td>102</td>
</tr>
<tr>
<td>28. Relationship of Various Factors to Listening Test Scores in School B (Coefficients of Correlation)</td>
<td>102</td>
</tr>
<tr>
<td>29. Relationship of Other Factors to Listening Test Scores with Chronological Age Scores Held Constant, School B (Partial Correlation Coefficients)</td>
<td>103</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>30. Summary: Relationship Between Listening Test Scores and Various Predictors: School B</td>
<td>104</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

It is generally agreed that the enjoyment of music by the listener and performer alike is enhanced by an intelligent understanding of it. A purely sensory response may bring pleasure, but a complete response includes both sensory and intellectual responses interacting in the listener. To enable children to gain pleasure and to quicken their sensitivity to the rich heritage of the music of the world, music educators are striving to develop music curricula from which children will grow in total sensitivity towards music—not only affectively, but cognitively as well.

Problem

The Conceptual Approach to Teaching Music

To provide such a curriculum, music educators in the past decade have emphasized the importance of teaching the basic structural concepts of music. Bruner has stated:

Grasping the structure of a subject is understanding it in a way that permits many other things to be related to it meaningfully. To learn structure, in short, is to learn how things are related.1

Identifying the structural elements of music, and basing a curriculum upon the development of such conceptual understandings thus becomes.

... not ... simply one of several possible things of interest to an educator; it is rather the essence of the concerns of an educator.¹

The Need for Experimental Work in Concept Development

There exist few instruments for measuring the identification of musical concepts. Andrews and Diehl have pointed out that most past tests have dealt with areas of music aptitude, music achievement, and music appreciation, but none has explored the area of musical concepts which are essentially cognitive in nature.² These authors, attempting to meet this need, developed a technique, the Battery of Musical Concept Measures, which seemed to identify various degrees of musical concept development in fourth-grade children. One of the measures, the Listening Measure, sought to identify musical concepts through aural experience with standard orchestral literature. Since listening constitutes the chief musical activity of the majority of adults, teaching children to listen intelligently becomes


a project first in importance in any consideration of music.

The Listening Test constructed for this study sought to measure conceptual understandings of children through aural experience with such literature. The Andrews and Diehl technique was followed in the construction and development of the test. To find "where the child is," musically, in terms of conceptual development, must be the starting point for any curricular planning.

The Role of Environment in Conceptual Development

Equally important to the need for developing curricula based upon the basic concepts of a discipline, however, is the need for understanding the forces which inhibit or facilitate learning. Many recent articles, books, and research efforts concerning learning have dealt with the influences of environment upon children's ability to learn in school. It has been recognized that many children's inability to cope effectively with the formal education of the school is not due entirely to their ability or intelligence, but rather to factors related to their environment. Children from disadvantaged areas seem particularly limited.

. . . the poor—be they white, black, Mexican-American, or Puerto Rican—bring their environment with them into the schools. Society's sickness touches every subject in the curriculum, including music. The strain on every subject has been severe. It is breaking the backbone of many
city music programs. Experienced music teachers are leaving the profession or fleeing to the safety of the suburbs.¹

In the literature dealing with the lower-class child, two major problems of learning which would seem to have implications for the music educator have been emphasized. The first is the problem of an inability to listen carefully and discriminatively. Trubowitz reported that many children in ghetto elementary schools reveal this inability to listen, and he states that research confirms this as a major area of weakness.² Raph, in discussing the research studies of Cynthia Deutsch, also cites this difficulty. The Deutsch studies reveal that children from lower-class groups can be relatively inattentive to auditory stimuli, and appear to have difficulty with any skill that is primarily or importantly dependent on good auditory discrimination. Such difficulty results from insensitivity in listening habits developed as a result of the very noisy home environment.³ Since discriminative listening is the perceptual basis from which music learning proceeds, these


findings would seem to have important implications for the music educator.

A second difficulty which the lower-class child is said to experience is that of adequate verbalization. Again the home environment has been cited as one of the important factors. Raph stated:

Discrimination learning, wherein a child acquires and uses refinements of concepts ... is necessary to basic language development also, and it is susceptible to delay, if not conspicuous lack of development, in the deprived social circumstances of the lower-class child.¹

Retarded verbal ability appears to be one of the most basic handicaps of the lower-class child. In discussing studies done with children raised in institutions and orphanages, Raph reported that their greatest weaknesses were in the areas of concept formation, reasoning, and abstract thinking.

As late as adolescence, the institution-reared group had greater difficulty with tests involving learning songs, rhymes, and stories, recalling the past clearly, or anticipating the future ...²

Riessman points out, however, that the verbal inadequacies of the lower-class child may not be simply stated as verbal inadequacies as such, but as inadequacies with the language of the school. He states that these children appear to be poor in the use of verbs, but much better with

¹Ibid., p. 189.
²Ibid., p. 185.
descriptive adjectives.

Walter Murray reminds us that in everyday conversation deprived individuals demonstrate a language that is often rich in simile and analogy. . . . A large proportion of the new words that have become a part of our language (e.g., "oomph") are said to have their origins among deprived groups. Some of the words come in via musicians, while others come in through the hipsters and the Beatniks who have been much influenced by the culture of the underprivileged.1

Since the study of music does require a rather formal language, and since effective learning in music must depend upon communication between the student and the teacher using this language, these findings should also raise some questions for the music educator as to the most effective way to communicate with these children in music classes. Haddad has stated:

Because the child often has never been outside his own area nor had association with people other than his own, there is a problem of language to overcome. The ghetto child is very capable of communicating. He is able to express himself well about something with which he is familiar. The language used may not be the language of the masses. . . . There is evidence that what the inner-city child knows and what is expected of him in the school are unrelated.2

Thus, in general education, it has been shown that the culturally deprived child brings severe handicaps to his school learning situation. Is this true also of music


learning situations? Inner-city children often display a most comprehensive understanding of auditory stimuli in musical performance. Much of the music which is popular with youngsters today reflects the folk heritage, the "blues" tradition, the earthy quality derived from the early artists whose roots were in the lower class.

. . . The inner-city child is not musically deprived; he is well acquainted with music. His music is relevant and very personal; he understands it. The problem develops when the child's music does not conform to the musical culture of the teacher. Any music that can be made meaningful and relevant to his life will become his music.1

The Need for Experimental Work

Considering the Role of Environment

Is the child of middle-class background any more advantaged in listening perception and verbalization about music? Does his relatively advantaged background help him to communicate about and deal more effectively with musical concepts?

There appear to be few studies dealing with socio-economic background as a factor in listening perception or music verbalization by children. Studies have revealed that sensitivity to rhythmic and melodic movement, aurally received, follow the pattern of human growth and maturation closely.

No recent studies investigating the relationship between musical concept development and environmental backgrounds of children were found. The Schneider-Cady Evaluation and Synthesis of Research Studies Relating to Music Education cites an early study by Wheeler and Wheeler (1933). These investigators reported the musical aptitude of mountain children in East Tennessee to be below the Seashore norms when compared to other subjects, but as a group, their scores were above those found by ten other investigators using nonmountain children as subjects. A study by Shephard in 1942 which used the Kwalwasser-Dykema Test with 104 pairs of subjects reported no statistically significant differences in verbalization between urban and rural children.1

Gordon, in a paper dealing with a comparison of the performances of culturally disadvantaged junior-high students and culturally heterogeneous students, reported that the differences between the mean scores for the two groups were negligible and not significant. The largest standard score difference found was for the Musical Sensitivity Style test (2.6) which favored the musically select

culturally heterogeneous group. The Gordon Musical Aptitude Profile was the instrument used.

Laverty, in a discussion of the Andrews-Diehl Battery of Musical Concept Measures, reported that the location of the school in a predominantly urban, suburban, or rural area did not account for between school differences in the scores from the Verbal and Listening Measures.

The importance of music for all children is generally recognized by all educators. Federal enrichment programs have given emphasis to the importance of music as a part of every child's total educative experience. Whether music remains a recreational outlet for the child or becomes a meaningful form of communication with others and with himself perhaps depends upon a new look at the ways in which music educators seek to "educate musically." An inner city school principal has bluntly termed music and art as two of the best examples of the complete irrelevancy of American education:

In other words, in order to be educated you must like symphonic music, you must possess certain tastes. This is ridiculous. There are children in this inner-city junior high school who sit for

---


forty-five minutes being lectured to on the life of Beethoven. It's enough to turn your stomach. Clearly, a crisis in music education is at hand.

If no differences in basic conceptual understandings of music exist between culturally deprived children and any other children, then inferior programs of music should not be accounted for by the disadvantaged backgrounds of the students involved. Rather it is up to the profession to examine content, methodology, and procedure for improvement in programs. The challenge is to provide a curriculum that achieves successful participation on the part of all children, and to find those instructional materials and methods from which all children can learn. Environment's role must be evaluated and explored. It was for such exploration and evaluation that the Listening Test developed for this study was used in a pilot study of two school populations in contrasting socioeconomic settings.

In summary, the need for this study seemed to be twofold: a need for the further development of concept measures in music, and a need for research to understand the effects of environment on the learner's musical concepts.

Problem Specified

This study sought to determine whether or not significant differences existed between lower-class and

middle-class children's identification of musical concepts as measured by the Listening Test developed for this purpose. In addition, the study investigated the relationships between the factors of mental age, chronological age, vocabulary comprehension, and reading comprehension of middle-class and lower-class children and their ability to identify musical concepts within a frame of reference of standard orchestral literature.

**Purposes of the Study**

The purposes of this study were:

1. to determine whether or not differences existed between lower-class and middle-class children's identification of musical concepts found in standard orchestral literature.

2. to determine what relationships existed between the mental ages, chronological ages, vocabulary comprehension, and reading comprehension of middle-class children and their ability to identify musical concepts found in standard orchestral literature.

3. to determine what relationships existed between the mental ages, chronological ages, vocabulary comprehension, and reading comprehension of lower-class children and their ability to identify musical concepts found in standard orchestral literature.
Sub-Purpose of the Study

The sub-purpose of this study was to develop a listening tape of recorded orchestral musical excerpts for use in a Listening Test which would effectively measure differences or lack of differences in identification of musical concepts by children from contrasting socioeconomic environments.

Hypotheses

To investigate the main purpose, the following null hypotheses were formed:

1. There is no significant difference between lower-class and middle-class fourth grade children's identification of musical concepts found in standard orchestral literature.

2. There are no significant relationships between the factors of mental age, chronological age, vocabulary comprehension, and reading comprehension of middle-class fourth-grade children and their ability to identify musical concepts found in standard orchestral literature.

3. There are no significant relationships between the factors of mental age, chronological age, vocabulary comprehension, and reading comprehension of lower-class fourth-grade children and their ability to identify musical concepts found in standard orchestral literature.
Assumptions

The following assumptions were made:

1. It is possible to obtain the true identification of children's perception of musical concepts through the administration of the Listening Test developed for this study.

2. The school populations selected for the pilot study sufficiently represented contrasting socioeconomic backgrounds as to provide valid data for the study.

3. The fourth-grade sample populations were sufficiently representative of normal fourth-grade populations to allow reliable and valid norms for comparison.

Delimitations

This study included the fourth-grade populations of two schools in or near the city of Columbus, Ohio. No generalizations can be made concerning any other school populations.

Definitions

Musical concept: In this study a musical concept was determined to be the idea of a basic element of music. Included in the study were the elements of pitch, duration, and loudness as these functioned within a
frame of reference, namely, within orchestral literature.

Listening Test: This study used a Listening Test developed according to the technique described in the Andrews and Diehl study, *Development of a Technique for Identifying Elementary School Children’s Musical Concepts*.\(^1\) The Listening Test required written responses to musical stimuli aurally perceived.

\(^1\) Andrews and Diehl, *op. cit.*
CHAPTER II

REVIEW OF RELATED LITERATURE

The purpose of this chapter is to survey the literature related to concept formation in children from contrasting socioeconomic backgrounds, showing the theoretical development of the present study. The chapter will deal first with concept acquisition and formation in the learner. Second, studies related to concept formation and perceptual development of the lower-class child will be discussed. Finally, the chapter will review studies more directly related to concept formation with auditory stimuli, and will consider research directly related to musical listening perception and conception.

Concepts and Concept Development

Definitions

It is evident that all writers do not wholly agree on a definition of a concept, nor on the process by which concepts are acquired. Woodruff, in writing of concept teaching in music, defines concepts as mental images of things one has encountered and retained in his mind. He states that concepts are acquired only through the basic
senses, when they are in direct contact with real things.¹

In the MENC publication, The Study of Music in the Elementary School—A Conceptual Approach, he further develops this definition:

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 the person has seen. At its most abstract and complex level it is a synthesis of a number of conclusions he has drawn about his experience with particular things. A conceptual statement is a description of the properties of a process, structure, or quality, stated in a form which indicates what has to be demonstrated or portrayed so a learner can perceive the process, structure or quality for himself.²

From this same text comes this similar definition:

In this publication . . . the term concept will be used to indicate "that which remains in the mind following a given learning experience." . . . It may be a generalization or a very specific bit of learning that ultimately will be a part of a much broader concept.³

Concept Learning

Gagne, in his text, The Conditions of Learning, has set up a hierarchy of learning types from simplest to


³Ibid., p. 2.
complex. Concept learning is assigned the sixth type of eight types discussed. He describes concept learning as that type of learning which

... makes it possible for the individual to respond to things or events as a class. ... In a real sense, he classifies them.¹

Hunt sees concept learning as an experience "... in which a subject learns to make an identifying response to members of a set of not completely identical stimuli."² Viewed in this way, the process of making decisions becomes involved. Hunt states:

Concept learning is a decision-making situation. During learning hypotheses must be chosen. The use of a concept involves decisions. The learner must decide what name applies to a particular object.³

Hebb also cites this necessity to abstract consistency from sometimes varying stimuli:

... 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.⁴

Underwood recognizes that concept formation often requires stimulus selection.


³Ibid., p. 160.

From among the elements of compound stimuli the subject is required to select or identify one or more elements which are consistently present when another event occurs. ¹

Concept formation, viewed in this way, becomes a complex process governed by many elements. The learner's selection of stimuli to which to respond becomes an extremely subjective and important process—one which is governed by his past experiences. It would seem then, that concept formation is intimately linked with perception, or perceptual experiences.

Perception and Conception

The differences between perception and conception are sometimes vague and not clearly defined. Hebb states that percept and concept are intimately related; in describing how learning capacity changes with growth, he states that growth is essentially a conceptual development—

... Perceptual organization is also involved, but percept and concept are intimately related, and the term "conceptual development" will do to cover both.²

To Hunt, the differences between perception and conception are quantitative rather than qualitative.³


³Hunt, op. cit., p. 5.
Hunt feels that the real differences between conception and perception lie in the process involved. He states that, in terms of Gestalt psychology, the "set" with which the subject approaches the task is different in perception from that with which he approaches the concept-learning task.

The perceptual act of categorizing is immediate . . . During the learning of a perceptual classification, a stable analytic rule is seldom developed. Instead, the categorizer appears to learn to make the appropriate response without analysis of the component parts of the stimulus. In a concept-learning situation, . . . a relatively greater emphasis is placed on such an analysis. The concept learner is pictured as viewing a stimulus as an aggregate of examples of more molecular concepts and the relations between them.\(^1\)

According to Hunt, perceptual categorization precedes conceptual categorization.

In what we normally call perceptual pattern recognition, the information which is transmitted directly from the pattern to be categorized through the sensory receptors of the categorizer may not be sufficient to classify the pattern.\(^2\)

He cautions, however, that neither the process nor situational differences between perception and conception are "all or none" distinctions.

This view of perception as the behavior which precedes concept formation is held by a number of writers. Hebb seems to include this sequential organization in his

\(^1\)Ibid., p. 5.
\(^2\)Ibid., p. 5.
description of conceptual development. He states that con­ceptual development, in the very first stages, depends upon earlier experience to establish perceptual elements.

These are the entities that make up more complex per­ceptions. Organizing such elements in the various sense modes would lay the foundation of all later responses of the environment. Secondly, there is a period of establishing simple associations, and with them conceptual sequences—the period in which mean­ing first begins to appear. Finally, the learning characteristics of the mature animal makes it appear­ance. . . . This later learning is essentially con­ceptual.i

Woodruff clearly states this sequential process:

Learning goes through three main phases to produce a clear and useful concept. The first phase is per­ceptual, which means that the sense organs are collect­ing stimuli from the actual subject and are transmitting them to the brain in the form of bits of meaning. The next phase is conceptual, which means that the brain organizes the bits of meaning so they fit together in a form which is as near as possible to the form of the real subject. The cul­minating phase is applicatory, or experimental, or whatever one cares to call the use of an idea in behavior.2

Thus, for concept development to proceed, the act of per­ceiving, in terms of subjective experience, is extremely important. Lack of experience, or lack of the meaning in the experience would seem to inhibit conceptual development.

The Use of Symbols in Concept Learning

The capacity to use symbols seems crucial in concept formation. To Mowrer, concept formation is a matter of

1Hebb, op. cit., p. 117.
2Woodruff, op. cit., p. 221.
mediated, rather than immediate, generalization, involving the use of verbal symbols.¹ Hunt states that

The goal in concept learning is the attainment of a definition of a concept which provides a satisfactory decision rule for assigning names to objects.² While the role of verbalization, or "labeling" is not viewed as an intrinsic part of the acquisition of concepts by all writers, the ultimate usefulness of verbalization is generally emphasized. Gagne describes the role of the verbal associate as an "important shortcut" to the learning of a concept.

The function of the word is that of an external stimulus to recall, which makes it relatively easy to structure the situation required for learning.³

Cofer, in a paper dealing with the acquisition of syntax, states,

Labelling may not come until very late, if at all, in concept formation; but if it is introduced, it often facilitates concept formation.⁴

Hebb does not connect language or verbalization so closely with conceptual development as do some writers. He states that (a child's) response to the verbal cue, as with (for example) "right" and "left," may remain vacillating and


²Hunt, op. cit., p. 160.

³Gagne, op. cit., p. 131.

unpredictable for some time, even though the general concept of "sidedness" may already show considerable conceptual elaboration as having taken place.¹ He would, rather, differentiate concepts as being "verbal" only, or "non-verbal."

Despite reliable and predictable verbal responses in the recognition of emotion, for example, a subject may be quite incapable of saying what considerations determined his choice of terms. There may, consequently, be concepts in man that do not have a verbal element, just as there are in animals. On the other hand, it seems quite clear that many concepts are fundamentally and essentially verbal—the "core" is a word or other symbol, without which the concept could no longer be an element in thought.²

Woodruff describes the verbal activity as the means for referring to our concepts. He also states that one needs language for formulating sharp and definite concepts from "relatively fuzzy impressions."³ Hunt describes Piaget's "concept" as an explanatory rule or law, by which a relation between two or more events may be described. Such explanatory statements need not be classification rules, nor may they say anything about the assignment of a name to a particular object.⁴

The verbalization of concepts is generally thought to be very important in the education of children, not only for

¹Hebb, op. cit., p. 118.
²Ibid., p. 133.
³Woodruff, op. cit., p. 219.
⁴Hunt, op. cit., p. 8.
the facilitation of the conceptual process of learning, but for the opportunity of communicating with the child about the meaning of his experiences. It is also important as a means of helping the child in stimulus selection.

Language has been described as "... a socially conditioned relationship between the child's internal and external worlds."\(^1\) When a child can successfully use words as mediators, he can "... effectively change his own social and material reality."\(^2\)

Summary

In summation, concept learning is dependent upon the experiences of the learner, and upon the analytic perception of both present and past experiences. Though a highly individualized process, it can be facilitated by the presentation of meaningful experiences, guidance in the analysis of sensory data, and help in the application of a satisfactory label. Though dependent upon maturation, it can be affected also by richness or deprivation of experience, and by fluency in the use of language. Thus it would seem quite irrevocably connected with environment.


\(^2\)Ibid.
Concept Learning and the Lower-Class Child

The lower-class child has been described as being different from the middle-class child in many ways. The recent decade has seen a veritable storehouse of literature dealing with this child, his background and environment, and his difficulties in school learning situations. Pertinent to this study seem to be the studies dealing with his auditory acuity, his capacity for using language, and to a lesser degree, the attitudes and value judgments he brings to a learning situation. The literature reviewed here is admittedly selective, and will deal with those studies and writings which seem especially relevant to this study.

Auditory Acuity

Since concept formation, in the early stages, is highly dependent upon sensory experience, and since the sensory data in music learning would seem to be acquired most logically through listening, the first part of this review will deal with auditory aspects of the lower-class child.

There are a number of studies which have differentiated between disadvantaged and nondisadvantaged children in their ability to discriminate between speech sounds. Recently, several studies have found differences which suggest difficulty with a much simpler auditory task, the detection of pure tones.
A recent U.S. Office of Education project, reported by Goldman, has dealt with cultural factors and hearing ability. The subjects in this study were 226 recent high school graduates selected by Fisk University as having the intellectual potential for successful academic achievement at the college level. All of these students were from a culturally and economically limited background. More than 10 per cent, or a total of twenty-five students failed the initial hearing evaluation test and were found to be unable to respond to pure tones at suprathreshold levels in the screen test situation. A retest showed that only one student suffered an actual hearing loss. The remaining twenty-four students were later found to respond to 0 Db. hearing level when alone in a sound isolated test room.¹

One can speculate with these investigators that the limited verbal interaction in the presence of high noise levels frequently found in the disadvantaged home environment might interfere with the acquisition by the child of the ability to extract an auditory signal from a competing background.²

Reported in the same article were the results of a recent school hearing survey in Sarasota County, Florida. In the report of the data compiled by McAdoo, it was found that, of the 3,634 second, fourth, and sixth grade pupils evaluated, 12 per cent failed the initial pure tone screen test. Although this percentage is somewhat higher than


²Ibid., p. 490.
that usually found in a school survey, the author states that it is certainly not phenomenal. A breakdown of the schools according to area, however, produced some "startling results."

In only those schools in nondisadvantaged areas the percentage of screen test failure was only 5.6 whereas the percentage of failure in the disadvantaged neighborhood schools was 41.4. Furthermore, on followup testing, almost all the latter group were found to have normal hearing.¹

Goldman suggests that the variable in both surveys was a significant disability in responding to a simple stimulus in a distraction situation. He poses the question, "What is the relationship between cultural background and auditory perception?"²

A study dealing with auditory perception of speech sounds is reported by Clark and Richards. The subjects were "Headstart" children attending the laboratory school of the University of Wisconsin. The instrument for measurement was the Wepman Auditory Discrimination Test. These investigators found a significant auditory deficiency in the economically disadvantaged population. A comparison of children on the Wepman Test revealed the disadvantaged children to be significantly poorer in the ability to differentiate between phonemically similar words.³

¹Ibid.
²Ibid.
The studies of Cynthia and Martin Deutsch in relation to the lower-class child are extensive and well-known. In a report relating auditory discrimination and learning, Cynthia Deutsch emphasizes the importance of experience:

However, for discrimination and recognition, another set of variables is as crucial as the intactness of the brain. These variables have to do with experience and exposure to the adequate stimuli. Thus, one might have fully adequate sensory apparatus, both peripheral and central, and yet, on being exposed to a spoken foreign language for the first time, be unable to recognize any referents to or meanings of the words heard—perhaps also finding it difficult ... to discriminate sounds or words from each other. These discriminations come only with experience and practice in responding to the stimuli.\textsuperscript{1}

The data reported on in the Deutsch study were gathered in connection with several different projects. All samples reported had in common data on auditory discrimination gathered through use of the Wepman Auditory Discrimination Test. Subjects were all lower-class children identified as poor readers or good readers at first, third, and fifth grade levels.

Correlations were reported between the Wepman Test and other measures, including the Lorge-Thorndike Test, a non-verbal intelligence measure at this level, and the Peabody Picture Vocabulary Test, which is an intelligence measure using verbal behavior. It was found that the Wepman

correlated with verbal measures, but not with non-verbal or performance ones, pointing up the relationship between auditory discrimination and verbalization.

The data presented were summed up in terms of the characteristics of poor readers: they have more difficulty with auditory discrimination; they have greater difficulty in shifting from one modality to another and back again; and they are more inefficient at a serial learning task when the stimuli are auditory as compared to learning when the stimuli are visual.¹

**Language Skills**

The lower-class child has also been found to have more difficulty with language skills when compared to his middle-class peer. Since communication dealing with musical concepts does involve a rather precise use of syntax, some studies relating to language skills are included in this review.

At the Institute for Developmental Studies, Martin Deutsch and associates conducted a study, the "Verbal Survey," with regard to possible interrelationships among language and some demographic variables. In the series of studies the attempt was made to identify patterns in the context of background variables at two developmental stages—the first grade and the fifth grade—and to relate these

¹Ibid.
background patterns to specific cognitive and linguistic patterns.

The population included a core sample of 292 children and an extended population of about 2500 children of various racial and social-class groupings. Negro and white, lower- and middle-class children were included.

Findings indicated that lower-class children, Negro and white, when compared with middle-class children, are subject to what the authors have labeled a "cumulative deficit phenomenon," which takes place between the first and fifth grade levels. Regarding the language ordering skills, it was found that as the complexity of the levels increased, from labeling, through relating, to categorizing, the negative effects of social disadvantage are increased.\(^1\)

It is also true, in looking at the enumeration scores, that as labeling requirements become more complex and related to more diverse and variegated experience, lower-class people with more restricted experience are going to have more difficulty in supplying the correct labels.\(^2\)

Robinson and Nukerji reported on an exploratory study undertaken in the Spring 1964 term in a kindergarten class in a "special service" public school in Brooklyn. The ethnic groups represented were Negro, Puerto Rican, Jewish, Greek, Italian, and Chinese. The study was specifically


\(^2\)Ibid., p. 86.
concerned with guiding children toward verbal forms of communication, generalization, and logical thinking.¹

These children lacked fundamental concepts of conceptualization itself, as revealed in their inability to classify or categorize items in their "store" in the schoolroom either in conversation or in fact. By working with the children, through "play" situations, the teachers sought to make them aware of the function of symbolic representation.

There were many examples in the exploratory study of the children's ignorance of various symbols or of the meanings or uses of symbols, which appeared to restrict their ability to gather information or to process it adequately.²

Gradually, an increasing number of children tried to classify items at "clean-up" time, and used classification terms with greater frequency and accuracy.

But teacher intervention was clearly needed with this class, in contrast to findings of a study in which middle-class kindergarten children discovered for themselves.³

In a discussion of a study using the Peabody Picture Vocabulary Test as a part of a larger investigation, John and Goldstein reported that the lower-class child's difficulty with certain words lies not in the frequency of the

²Ibid., p. 141.
³Ibid.
word, as found in one child's experience contrasted with a child from a differing socioeconomic environment, but in the opportunity the child has to engage in active dialogue using learning labels with an adult who supplies corrective feedback.

The larger study was designed to gather information about verbal skills, intellectual performance, and motivational approaches of white and Negro, first- and fifth-graders, lower- and middle-class children in New York City. A Concept Sorting Task was used to investigate category information in young children. The results of this study showed that

... the middle-class children tended to produce category labels more often than their lower-class peers, who were instead inclined to focus on non-essential attributes.¹

The lack of corrective feedback in the lower-class child's environment was suggested as a possible reason. The importance of corrective feedback and opportunity to use language as a tool for labeling and ordering stimuli in the environment is also brought out in a study by Shipman and Hess.²

Though the studies discussed thus far have not been related to music listening and verbalization, one is led to

¹John and Goldstein, op. cit., p. 171.
wonder whether or not the same lack of ability to cate-
gorize, classify, and "label" musical events might result
from the lack of opportunity to hear serious orchestral
music, such as is presented in children's listening lessons
in school music situations. Perhaps, too, the lack of any
guidance in "labeling" these sounds might produce similar
lack of success in music learning situations. One may
question, however, whether or not the middle-class child,
while perhaps having more opportunity for exposure to the
sound of classical orchestral music, has any more oppor-
tunity for guidance in labeling—"corrective feedback"—than
does his lower-class peer. John has stated that:

The middle-class children seem to have mastered the
skill of choosing the most appropriate single
response when presented with a complex task, while
their behavior is similar to that of their poorer
age-mates when they are required to enumerate and
describe.1

The above observation was made as a result of data collected
as part of a program, "Verbal Enrichment and School Achieve-
ment in Lower Socioeconomic Groups."

The search of literature revealed one study which
had as its purpose the analyzation and comparison of verbal
and musical listening abilities in elementary school chil-
dren. Wilson, testing 369 sixth-grade pupils, administered
the Form 4A Listening Test of the Sequential Tests of

1Vera P. John, "The Intellectual Development of Slum
Children: Some Preliminary Findings," American Journal of
Educational Progress, the Gaston Test of Musicality, the Whistler-Thorpe Musical Aptitude Test and the Kyme Test of Aesthetic Judgment. He found that there was a relationship between skill in music listening and language listening, perhaps because of some common factor of a general nature which contributes to skill in both. He felt, however, that existing tests were not reliable enough to measure such relationships definitively.¹

A number of subjective observations were found in the literature concerning the lower-class child's affective and motor response to music. Pavenstedt wrote in her study of low, lower-class children:

An outstanding trait was their great sense of rhythm. As with autistic children, they could be reached by music, and could much more readily memorize a verse when it was set to music. They listened attentively and responded with rhythmic body movements.²

It would have been interesting to know how these children would have "verbalized" what seemed motorically instinctive. What is the relationship between performance and verbal identification of musical sounds by children? Mainwaring found little apparent relation between the


ability to respond physically to a rhythmic stimulus and the ability to perceive cognitively the rhythmic pattern. Bond seemed to find no statistically verifiable relationship between rhythmic perception and motor performance of selected sports skills taught in physical education. The Seashore Test of Rhythm was used.

Lloyd gave this report of his experiences with lower-class children:

Youngsters can't get enough of "singing time." . . . They sing folk songs . . . with feeling and love. Although they can clap and sing complicated rhythm patterns . . . with much pleasure, they love simple songs. . . . They respond enthusiastically to classical music and quickly recognize the differences in movement and feeling. . . .

Motivation and Attitude

In the literature dealing with the lower-class child, much attention is given to the problem of motivation. Inextricably linked with the lower-class child's lack of motivation in school learning situations is his low self-concept, the "failure syndrome," and the lack of relevance that the school experiences seem to have for him. Davis has described the reaction of the lower-class child to "test"

---


situations which have little relationship to his experience:

[Any] unrealistic problem can arouse the child's desire to achieve a solution only if the child has been trained to evaluate highly any and all success in tests. No matter how unreal and purposeless the problem may seem the average child in a high socioeconomic group will work hard to solve it, if his parents, his teacher, or other school officers expect him to try hard. The average slum child, however, will usually react negatively to any school test, and especially to a test whose problems have no relation to his experience.¹

Therefore the cultural motivation of the extreme socioeconomic groups to succeed in a testing situation would seem to be very unequal.

Perhaps another aspect of motivation would involve preference for one type of music to another. Baumann, in investigating the musical preferences for three age groups --12-14, 15-17, and 18-20---found that the music items most preferred were in popular, rural, and western idioms. All subjects, regardless of socioeconomic background, disliked classical music more than any other type presented.

Children of the same grade levels in different sections of a city varied significantly in marking their reactions upon hearing the excerpts in the Music Preference Inventory which had been taken from popular, traditional, and classical music. . . . Sectional differences appeared when comparing Maryland and Arizona schools. The eastern schools preferred traditional music more while the western schools preferred classical music more. Young people of

extremely low income families indicated little preference for any kind of music.1 (italics mine.)

If motivation is important in listening perception, the findings described would perhaps indicate that preference for a type of music could perhaps affect the auditory acuity of children. They would attend to what they liked and were familiar with, and would perhaps "turn off" what they did not like. Socioeconomic position often serves to bring about familiarity with certain types of music and tends to exclude other types from the realm of experience.

A study by Fisher also involving musical preferences was conducted to determine how various age and economic status groups would react to classical type musical compositions whose identity was unknown to them. A limiting factor in this study was the single type of musical composition presented as stimulus. All the compositions were defined as "classical." Two hundred and fifty-one subjects, ranging in age from 10 to 25 were tested. They were asked to rank their preference for five classical recorded orchestral compositions. Agreement in preference predominated in the comparison of age groups, and no clear-cut preference differences were found between age groups,

socioeconomic groups, or sex groups.¹

Yet the question of motivation remains important in choosing methods and materials for the lower-class child. Andrews calls for the teacher of such children to be willing to give up mass application of musical materials that have long constituted standard curriculum content, and give up the assumption that the child will like the music the teacher presents.

Obviously, then, new ways must be found to engage the interest and attention of such children in the arts, or the arts will price themselves out of the endeavor to produce a culturally knowledgeable public, simply on the basis of lack of interest, or even positive rejection. . . . It . . . calls for music educators who understand the language, standards, background, experiences, and probable reactions to those who are taught.²

A study by Rogers would seem to give some indication of preference for type of music with different grade levels. A test using 57 pairs of musical excerpts was administered to 635 fourth, seventh, ninth, and twelfth-grade students. Not only was age found to be important concerning preferences, but also socioeconomic class. Though popular music was chosen to a greater and greater degree with increased age, socioeconomic status was a strong enough factor to


cause a difference in the preferences, as shown by the consistently larger number of choices made in favor of classical music by the upper class group.¹

Musical Concepts and Listening Perception

The concern for a music curriculum based on conceptual development is fairly recent. A search of the literature reveals few studies which are identifiable as investigations in conceptual learning in music. The studies dealing with listening perception, however, seem to bear relationship to this area. This section will be concerned with studies investigating differences found in musical discrimination with regard to socioeconomic background, or to differences in amount of musical experience. Since past experience has been shown in other areas to be a significant factor in tests of auditory discrimination, and since experience has been linked with socioeconomic level, studies dealing with this differentiation have been included in this section of the discussion of related literature.

Effect of Socioeconomic Status

Andrews and Diehl sought to identify children's basic concepts of music in their study, Development of a Technique for Identifying Elementary School Children's

¹Vincent Robert Rogers, "Children's Expressed Musical Preferences at Selected Grade Levels" (unpublished doctoral dissertation, Syracuse University, 1956).
Musical Concepts. Four measures were developed for testing musical dimensions of pitch, duration, and loudness. One of the measures was a Listening Measure which used musical excerpts from orchestral literature as stimuli. In the development of the final form of the Listening Measure, some evidence of differences in mean scores perhaps attributable to general cultural background was noted. In the trial designated as the University Area Trial, test items were administered to twenty volunteers from communities near the Pennsylvania State University. They were about equally divided between children who would soon enter and those who had recently completed fourth grade. The measure in this trial consisted of twenty-four items. The results of this trial showed a mean of 17.4, the maximum score being 24. Since the mean was considered high, a second trial was made with a group of 32 fourth-grade pupils, described as a "non-volunteer" group, from a different community. The obtained mean score for this trial was 9.50.

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 Estrella 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.¹

¹Andrews and Diehl, op. cit., p. 22.
Subsequent trials included both urban and rural fourth grade classrooms to insure the measure's reliability for a heterogeneous fourth-grade population.

Gordon, in a paper dealing with the performance of culturally disadvantaged students on his Musical Aptitude Profile, concluded that there were no significant differences between means for a socially heterogeneous group and the disadvantaged group of seventh-grade students who were enrolled in two junior high schools in a large north central city.

The MAP standard score scale has a range from approximately 20 through 80 based on a standard deviation of 10 for the total grade range 4-12. A comparison of mean scores obtained for the culturally disadvantaged groups with those obtained from the culturally heterogeneous groups shows that nine of the eleven differences were only one standard point or less for the musically unselected, as were seven of the eleven differences for the musically select. That these differences are negligible and have no practical significance is particularly emphasized by the fact that composite test mean standard score differences were .7 and .4 for the musically unselected and musically selected groups, respectively.¹

In the discussion of the apparent lack of difference in musical aptitude between the two groups, it was noted that

... even with a high degree of self-selection (one out of ten) in the two schools, there are, unfortunately, many culturally disadvantaged musically talented students who are not taking part in school music activities.²

¹Gordon, op. cit., p. 261.
²Ibid., p. 262.
Of the 55 students in the disadvantaged group who scored above the 90th percentile, only 13 students were in school music activities. Gordon speculated that perhaps general environmental factors could prevent a culturally disadvantaged musically talented student from achieving according to his potential. In the school of the lower-class child, perhaps, music study does not always attract the talented.

**Effect of Experience**

The studies of concept development cited in the first section of this chapter commonly emphasized the significance of the amount of experience in the development of discriminatory listening, and, consequently, in conceptual development. Yet, regarding music, the available research seems to indicate that previous experience seems generally to have little effect upon discrimination or upon preference. Rubin studied the relationship between musical experience and musical preferences and technical discrimination. His subjects were 100 students from the seventh, ninth, and twelfth grades. Fifty of these had scored the highest and 50 the lowest on a questionnaire designed to determine the amount of the subjects' musical experience. Care was taken to include in the sampling schools having students of differing socioeconomic background "... as socioeconomic
background is somewhat related to opportunity for musical experience."\(^1\)

In Rubin's test of discriminatory ability, the subjects were asked to compare pairs of associated musical examples, and to indicate whether or not any alteration had taken place in the second one. They were also asked whether the alteration was harmonic, melodic, or rhythmic. Art, folk, and "current vogue" selections comprised the 45 item test. Both groups achieved the highest percentage of correct responses in alterations of "current vogue" selections, but perceiving the nature of the alteration proved more difficult than recognizing that an alteration had taken place.

Rubin concluded that musical experience has little effect on preference, or on discriminatory ability, and, furthermore, that musical experience has little effect on the ability to identify relative differences in musical factors such as harmony, melody, and rhythm. All subjects found rhythmic discrimination the easiest, melodic discrimination comparatively simple, and harmonic discrimination the most difficult.\(^2\) If discriminatory skill and well-rounded musical interest are desirable goals in a school


\(^{2}\)Ibid.
music program aiming toward increasing the students' general capacity for appreciation, then music educators must find new ways for implementing these skills.

**Effect of Training**

Gernet sought to provide objective data relative to the growth and character of music appreciation in his study dealing with musical discrimination at various age and grade levels. The study's stated purposes were also to ascertain whether or not the principal aim of music education was being realized, and to determine the degree of deviation from accepted standards of musical literacy. To achieve these purposes, a Music Preference Test was constructed which was designed to test the capacity to make critical judgments of musical quality as exemplified in compositions of diverse merit. The Kwalwasser-Dyckema Musical Aptitude Test and school psychological tests also provided data. His findings were:

1) Aptitude is probably a more vital factor in the making of musical preferences than intelligence, but the predictive value is low.
2) There is a higher average relationship between intelligence and talent than between intelligence and preference.
3) There is a substantial and significant relationship between training and preference; musical training is the most closely related factor influencing the level of appreciation.
4) The general tendency for successive means of all ages is upward, and this characteristic is most clearly delineated in the two-year groupings.
5) Age is not highly correlated with preference; there was little appreciative change in the mean
preference scores for ages 12 to 17 inclusive, but above that range the average scores increase by comparatively large amounts.¹

An early study by Mainwaring seemed to emphasize the need for specific training for attaining discriminatory skill in musical situations. He states:

As the most complex forms of music can consist only in combinations of sounds "systematically differentiated" in pitch, intensity, duration and timbre, it follows that musical ability, being confined for its exercise to a world of sound differentia, must at least include, as a fundamental element in its complexity, the cognitive ability to educe, between sounds, these relations of difference.²

His subjects had an average age of 10.6 years. There were 23 girls and 29 boys. The boys had received no special musical training, and the girls had only a background of the standard "singing" in music classes.

His first discussion suggested by the results of the untrained group had to do with the general poverty of each of the abilities tested, but more especially the ability to identify rhythmic form.

The ability to perceive even the simplest rhythmic pattern cannot be assumed to have any general existence. . . . It would seem . . . that the apparently fairly physical reaction to regularly repeated sounds and any pleasure derivable therefrom and therein, do not necessarily involve the purely cognitive ability to perceive rhythmic pattern.³

¹Sterling Gernet, "Musical Discrimination at Various Age and Grade Levels" (unpublished doctoral dissertation, Teachers College, Temple University, 1939).
²Mainwaring, op. cit., p. 181.
³Ibid., p. 185.
He found that identification of pitch differences and perception of rhythmic patterns are not necessarily related skills, but both are at least in part related to the general factor in intelligence.

Mueller also reported this lack of ability to apprehend musical events and use the proper labels for manipulating and communicating the experience. An experiment in musical analysis was planned to measure as objectively as possible the musical qualities which could be observed by a wide variety of listeners who spent an hour listening to three repetitions of the third movement of Mozart's G Minor Symphony, no. 40, and answered questions about it. The subjects were 117 freshmen, sophomores, and juniors at Indiana University. Less than one-fourth of the subjects could identify the \( \frac{3}{4} \) meter; only half could identify the repetitive character of the melodic themes; only two questions on instrumentation were answered correctly. She stated that progress in apprehending music has been shown to be significantly related to formal training, interest, attitude, and to verbal intelligence. "There is previous evidence to indicate that it is also probably related to auditory sensitivity, and perhaps also related to measurable traits of temperament or intellect."

---

Zimmerman found that the extent of experience in performance groups had but little effect upon scores in a test developed to identify verbal-descriptive listening skills of high school students. Correlations of test scores and extent of experience in band, orchestra, chorus, private study, general music, music theory, and music appreciation ranged from .11 to .29. The highest relationships (.37 and .40) were found between test scores and total performance experience.1

Diehl, in an investigation of the relationships among concept development, listening achievement, musicality, and the amount of musical performance experience of college students found lack of significant relationships between concept development and performance experience, or concept development and listening achievement. He found a low but significant relationship between concept development and musicality, indicating that students with high musicality have not had meaningful instruction in musical concepts as measured in his study. Musicality also showed a low but significant relationship to scores on the listening achievement measure, indicating that the students

had not benefited in accord with their potential in the areas of listening achievement.¹

That listening skills can be taught successfully is indicated by the studies of Rasmussen, Graham, and Peterson. Rasmussen developed programmed tapes for teaching basic listening skills to fourth graders. He concluded that the skills of perceiving rhythm, melody, tone color, and aspects of texture and form could be taught through planned lessons.² Graham also used taped listening lessons for fourth and fifth graders, and found that listening ability improved.³ Peterson found that listening with notation produced significantly better results with seventh graders, but not in earlier grades.⁴


Difficulty of Verbalization

Difficulty in verbalizing, or providing "labels" for musical events has been found to exist rather universally. Laverty, in studying a socially heterogeneous population of third, fifth, and seventh graders, found that children frequently gave incorrect labels to musical changes which they seemed to perceive. The common use of "higher" and "lower" for changes in loudness as well as pitch was the most apparent confusion in labeling.¹ Riley and McKee reported that loudness discrimination was learned most easily by all subjects (grades one to three and adults), and that pitch was more difficult for children. They concluded that "... We have little doubt that 6-year-olds are more familiar with the concept 'louder-softer' than with 'higher-lower.'"² Pflederer, in a study designed to measure conservation in musical tasks, concluded that the conservation of meter and rhythm pattern is more difficult than conservation of tonal pattern, but that subjects in all age groups lacked the vocabulary needed to describe the musical stimuli.³

¹Laverty, op. cit.


Summary

The literature reviewed in the first section of this chapter, that dealing with the lower-class child, would perhaps indicate that the lower-class child can be differentiated from the middle-class child in terms of auditory acuity, manner of dealing with a school-learning situation, and value judgments of learning experiences. The search of the literature revealed no studies directly investigating these findings in regard to music learning situations. Hence, only assumptions can be made that a difference might also exist in those situations.

Contrarily, the studies dealing directly with music have indicated that, in the populations studied, few differences in listening perception or preference can be accounted for by previous experience or by socioeconomic level. Though some differences in preference and discriminatory listening have been observed, they cannot easily be generalized. Implications or predictive generalizations for the music educator seem to be that, considering the characteristics of the lower-class child, differences might be expected to exist, but we have little evidence that in music they do.

In attempting to summarize the research relating directly to music skills, perhaps some different conclusions might be speculated upon. The studies reviewed which related to music have not necessarily been confined to
those dealing with the musical development of children. Nor have they attempted to define the sequential stages of musical development. These studies have involved children, college-age students, and adults. However, they all involved heterogeneous social groups. It can perhaps be assumed that lower-, middle-, and upper-class subjects were included in all the samples. Therefore, if one should attempt to draw any conclusions simply on the basis of the music research reviewed, it would probably be that the problem of discriminatory listening, if viewed as a desirable goal of musical education, exists to the same degree in all children, regardless of background. Perhaps this is true. It should follow logically, then, that no greater problem exists for the music educator in the lower-class school than in any other school. Music educators should experience equal success, or lack of success, in lower-class schools as in any other. Yet, of course, this has not been true. The lower-class school music educator has experienced the same difficulties as have the teachers of other subjects in these schools in reaching the children and in making education a vital force in their lives. The music educator has realized that many of the music activities in the lower-class school have failed to develop an appreciation for or even an awareness of the richness of the world's musical heritage. Yet for the lower-class child, the development of enlarged musical horizons which
would enable him to use music as a means for self-expression, self-identity, or self-satisfaction seems particularly vital.

One important task, it would seem then, is to find first where children are, musically, at any age level, with environment as one important consideration. If it is found that environment does, in any way, differentiate children in terms of musical perception, listening skills, or general musical development, then a reassessment of the standard methods and materials would perhaps be necessary. If ability is equal, perhaps "readiness" is not. This study attempts to focus upon but one aspect of an admittedly larger problem. Through objective evaluation of an existing situation, the search for solutions can be begun.
CHAPTER III

THE DEVELOPMENT OF A TEST OF MUSICAL CONCEPTS

The procedures of the study included (1) the selection of musical excerpts from the orchestral literature which could be used in a test of children's concepts of change in duration, loudness, and pitch, (2) the development of a listening measure for testing these concepts, (3) the use of the listening measure in a pilot study of two socioeconomically differing school populations, and (4) the correlational treatment of (a) the data derived from testing two student populations with (b) the data obtained concerning the students' backgrounds and characteristics. Chapter III will describe the development of the test, and Chapter IV will describe the pilot study of the two schools.

Procedures

The development of a test which would show children's concepts of musical duration, loudness, and pitch was the first objective of the study. In order to fulfill this objective, specific procedures were followed as recommended in the test construction literature. These included (1) selection of a test format, (2) selection of test items, (3) pretesting the items, and (4) analyzing the items.
through item analysis. The activities undertaken in this phase of the study are described in the sections which follow.

Selection of a Test Format

The test of musical concepts developed for this study was modelled after the Andrews and Diehl Listening Measure described in detail in the study, Development of a Technique for Identifying Elementary School Children's Musical Concepts. In describing the Listening Measure developed in their study, these authors stated:

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.

Such identification would seem to require not only a cognitive judgment of a sensory experience (conceptual understanding), but a prerequisite selection of a stimulus from a pattern of stimuli. It was decided, therefore, that the Andrews and Diehl Listening Measure would provide a suitable model for the test to be constructed for this study.

Selection of Test Items

Recent music texts for elementary grades have shown increasing attention to "listening experiences" as part of

---


2 Ibid., p. 17.
the total music program. A great part of the musical literature toward which music educators are guided for these "listening lessons" is orchestral literature. Many of the song texts have such literature included in the recordings which accompany them. Other sources include the RCA Adventures in Music and the Bowmar Orchestral Library, two major "listening libraries" developed for use in the elementary school. These various sources were used in selecting the musical items for the test developed in this study. These items were believed to have content validity because experts in elementary school music had been the compilers of the various sources for the items.

Item Content

Following the technique described in the Andrews and Diehl study, three types of items were developed:

(1) In the first group of items (musical excerpts), a predominant change in one dimension occurred within the example, e.g., the music became slower, faster, louder, softer, higher, or lower.

(2) In the second group of items, predominant changes in two dimensions occurred within the example, e.g., the music became faster and louder, lower and slower, or some other combination.
(3) In the third group of items the excerpts were paired. Each pair of musical excerpts differed predominantly in one dimension. For example, the second excerpt might be higher than the first, though closely related in melodic content. The subjects were asked to indicate the change that had taken place in the second excerpt.

Each of the excerpts in the first two sections was repeated. The paired excerpts in the final section were not repeated.

In the construction of the test, a rigorous attempt was made to include examples in which the change could be perceived as the most important one. In order to validate the discernibility of these changes, the music teacher of each group tested was asked to complete a response sheet. Their responses were carefully examined for the purpose of determining content validity. Two validation trials with adults possessing a variety of musical experiences were also carried out for this purpose.

**Pretesting the Items**

In order to determine the feasibility of developing such a test using the orchestral literature available for children's listening experiences, a series of trial tests were developed and administered. Each of these trial tests is described in detail in subsequent sections of this chapter. The items used in the instrument for the pilot
study were those which had been found accumulatively useful as a result of item analysis information obtained in each trial.

**Item Analysis**

Item analysis is the application of statistical techniques to assess two characteristics of items—their difficulty and the extent to which they are correlated within the measure and with other measures. From the assessment of individual items emerges the concept of the reliability of the test, or the consistency with which it measures a trait in a particular group. The test construction literature provides guidelines and criteria for assessing the statistical data derived from item analysis. These were followed in the attempt to develop a satisfactory instrument for use in the pilot study.

The OSU Item Analysis Program developed and maintained by the Center for Measurement and Evaluation of the Ohio State University provided the statistical data for the analysis of the items. Description of these data and guidelines provided in the test construction literature are discussed in the following sections. A copy of the Item Analysis Program may be seen in Appendix D.

**Mean Score.**—For each trial test, a mean score was provided. This is the average of all scores obtained by summing the scores and dividing by the number of scores. The mean can provide some indication of the relative
difficulty of the test, and of the test's probable discriminating power.

**Standard Deviation.**—The standard deviation describes the spread or the variability of the scores around the mean and is a statistic associated with the normal curve. In a normal distribution, one standard deviation on either side of the mean encompasses approximately 68 percent of the cases.

**Range.**—Range is defined as the highest score minus the lowest score plus one. It can be used to determine the homogeneity of the group being tested. A limited range would indicate that the individuals being tested are very similar in reference to the trait measured. A wide range would indicate that the group is heterogeneous, the variability being great.

**Relative Difficulty.**—The relative difficulty index describes the difficulty of the item in terms of the percentage of students selecting an incorrect answer. As the percentage increases the item is more difficult. A difficulty index for each individual item was provided as well as a mean difficulty index for the total test.

Difficulty ranges may be interpreted as follows:

1. **0.00-0.20** — very easy items.
2. **0.21-0.40** — relatively easy items.
3. **0.41-0.60** — "average" items.
4. **0.61-0.80** — relatively difficult items.
5. **0.81-1.00** — very difficult items.
Downie and Heath have stated:

From the viewpoint of item difficulty, a well-made test starts with a few very easy items, continues with items of increasing difficulty, and ends with a few items which only a very few of the examinees answer correctly. There would be more difficulty values clustering about the center than at either extreme, but there would be a balance so that the average item difficulty is 50 per cent.\(^1\)

Nunnally\(^2\) has stated that there is, however, no certain exact difficulty level.

The \(p\) value is determined by both the intrinsic difficulty of the item and the effect of guessing. Guessing tends to make \(p\) values higher, the amount of "elevation" being inversely related to the number of alternative responses for each item. . . . To peak a test at .5 . . . would be inappropriate when there are four or five alternative responses, as typically found on multiple-choice tests.\(^3\)

He suggests that it is more sensible to construct tests primarily in terms of the actual correlation of items scores with total test scores.

**Item Discrimination.**—The OSU Item Analysis Program provided three indices of item discrimination: (1) a discriminatory index which reflected the degree to which the item discriminated between the upper and lower groups, (2) the correlational phi coefficient, and (3) the correlational point biserial \(r\) coefficient.


\(^3\)Ibid.
The discrimination index gives the ratio of the difference between the upper 27.5 per cent of the cases and the lower 27.5 per cent, unless otherwise specified. Davis has stated that items with discrimination indices above 20 will ordinarily be found to have sufficient discriminating power for use in most achievement and aptitude tests.¹ Many factors influence the item's discrimination index, including

... the item's clarity of expression and lack of ambiguity, the degree to which the keyed answer is adequate and incontestably correct, the number of items in the groups tried out, the intercorrelations of the items, the reliability of the item and criterion scores, the true relationship between the psychological functions measured by the item and by the criterion, the level of competence of the testees ..., and to a slight extent the level of difficulty of the item.²

The correlational phi coefficient is an item-item relationship between the upper and lower groups. It may be tested for significance. Downie and Heath have stated, "Because of the nature of the statistic, it favors items which have difficulty levels of 50 per cent."³ Items showing significance at the .05, .01, or .001 levels of confidence were regarded as the most useful items.


²Ibid., pp. 15, 16.

³Downie and Heath, op. cit., p. 198.
The point biserial r coefficient shows the relationship of the item to the total score on the test, thus giving a measure of the validity of that item. Nunnally has stated:

It is recommended that the PM [Product Moment] correlation be used in item analysis, which with dichotomous items is point-biserial. Not only does the PM correlation give very much the same information any other measure of item-total relationships would provide, but to the extent that item selection would be slightly different by different measures, the PM correlation is logically better than the other measures.1

Items with point biserial r coefficients significant at the .05, .01 and .001 levels of confidence were regarded as having the greatest validity. For reasons discussed later in the chapter, some items were included in the final form of the Listening Test which were not at the .05 level of confidence.

Reliability.—The reliability of a test, or the consistency with which it measures a trait in a certain group, is affected by many factors. Some of these are listed by Downie and Heath:

... The length of any test influences the size of the reliability coefficient for that instrument. Since reliability coefficients are correlation coefficients, they too are greatly affected by the range of scores in the sample on which the reliability correlation is computed. The more homogeneous the sample, the lower the reliability coefficient. The size of the reliability coefficient will differ when computations are based upon different samples. Thus no test has a single, characteristic reliability coefficient.2

1Nunnally, op. cit., p. 262.
2Downie and Heath, op. cit., p. 221.
It would seem, therefore, that a satisfactory reliability estimate would depend somewhat upon the ultimate use of the instrument. Nunnally has provided this criterion for assessment of reliability:

What a satisfactory level of reliability is depends on how a measure is being used. In the early stages of research on predictor tests or hypothesized measures of a construct, one saves time and energy by working with instruments that have only modest reliability, for which purpose reliabilities of .60 or .50 will suffice.¹

The obtained reliability estimate on the final form of the test developed for this study met this standard.

**Exploratory Trial, June, 1969**

**Purpose.**—To determine the feasibility of developing a test using the orchestral literature available and selected for children's listening experiences, a trial listening audio tape of eighteen musical excerpts was prepared. A copy of the sources of the musical items, in the sequence used on the tape, is presented in Appendix A. It is designated as Form A.

**Subjects.**—The subjects in this sample were 74 elementary education majors. They were enrolled in a course in music methods and materials for the elementary school, designated as Music 370 at the Ohio State University.

**Materials and Procedures.**—Eighteen items (musical excerpts) plus four sample items, in audio-taped form, were

used for this trial. The multiple choice format which was used in the Andrews and Diehl study provided the model for the response sheet. A copy of this response sheet is found in Appendix A. The subjects were instructed to listen to the example and mark the response describing the most important change.

Results.—In scoring this measure, each item received one point. A complete item analysis for the test was obtained from the Center for Measurement and Evaluation of the Ohio State University. The item analyses for the Exploratory Trial, Form A, are presented in Tables 1, 2, and 3. The scores for this trial ranged from 11 to 18, with a mean of 14.66 and a standard deviation of 1.58. The Kuder-Richardson "20" reliability estimate was .164 (see Table 1).

<table>
<thead>
<tr>
<th>No. Items</th>
<th>Range</th>
<th>Mean</th>
<th>S.D.</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>11-18*</td>
<td>14.66</td>
<td>1.58**</td>
<td>.164</td>
</tr>
</tbody>
</table>

*Total score possible = 18.
**N = 74.

The mean item difficulty was .185. This relatively low difficulty level would seem to indicate that the musical changes were quite easily discerned. Item 18, showing the highest difficulty level, was interrupted by the bell.
indicating the end of the class period, and hence cannot be regarded as showing a valid difficulty index. The majority of the scores fell in the "average" and "very easy" ranges. A summary of the item difficulty is given in Table 2. The item difficulty of each item is given in Table 3 with other data.

**TABLE 2**

**EXPLORATORY TRIAL; FORM A: SUMMARY ITEM DIFFICULTY**

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Mean</th>
<th>Range</th>
<th>Item Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>.00-.20</td>
<td>1,2,3,4,6,8,9,11,13,14,15</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.21-.40</td>
<td>5,7,10,12,17</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.185</td>
<td>.41-.60</td>
<td>16</td>
</tr>
<tr>
<td>1</td>
<td>.61-.80</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

The discrimination index in this trial reflected the ratio of the difference between the top 29.73 per cent and the lower 18.92 per cent of the group. Eight items in this trial had discrimination indices of over 20. Mean item discrimination index was 23.6.

Eight items had phi coefficients significant at the .01 level of confidence, and three at the .05 level. The point biserial r coefficients were significant at the .01 level of confidence for six items and for four items at the .05 level. Table 3 presents the complete item analysis and the dimension measured.
Two college instructors of the classes involved took the test at the same time the students did, and their response sheets were also examined for the purpose of determining content validity. Their responses indicated agreement with the investigator as to the musical change occurring.

Discussion.—The high mean, limited range, low difficulty level, and relatively large number of non-discriminating items indicated that this would not be a suitable measure for these college students. The purpose of this trial, however, was not to provide information about the instrument's power as a discriminatory test with these students, but rather to test the discernibility of the

### Table 3: Exploratory Trial; Form A: Item Analysis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.014</td>
<td>.324</td>
<td>.197</td>
<td>7.1</td>
<td>Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.189</td>
<td>.869*</td>
<td>.465*</td>
<td>57.1</td>
<td>Pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.014</td>
<td>.324</td>
<td>.271**</td>
<td>7.1</td>
<td>Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.014</td>
<td>-.203</td>
<td>-.099</td>
<td>-4.5</td>
<td>Pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.216</td>
<td>.707*</td>
<td>.303*</td>
<td>35.7</td>
<td>Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
<td>N.S.</td>
<td>Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.297</td>
<td>.661*</td>
<td>.235**</td>
<td>40.9</td>
<td>Loudness, Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.162</td>
<td>.869*</td>
<td>.533*</td>
<td>57.1</td>
<td>Pitch, Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.014</td>
<td>.324</td>
<td>.123</td>
<td>7.1</td>
<td>Loudness, Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.338</td>
<td>.673*</td>
<td>.336*</td>
<td>46.1</td>
<td>Pitch, Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>.014</td>
<td>.324</td>
<td>.123</td>
<td>7.1</td>
<td>Pitch, Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>.392</td>
<td>-.249</td>
<td>.039</td>
<td>-14.9</td>
<td>Pitch, Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>.122</td>
<td>.707*</td>
<td>.444*</td>
<td>35.7</td>
<td>Duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>.041</td>
<td>.454*</td>
<td>.303*</td>
<td>14.3</td>
<td>Pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>.162</td>
<td>.383**</td>
<td>.184</td>
<td>19.5</td>
<td>Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>.419</td>
<td>.853*</td>
<td>.495*</td>
<td>64.9</td>
<td>Pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>.230</td>
<td>.383**</td>
<td>.229**</td>
<td>19.5</td>
<td>Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>.703</td>
<td>.368**</td>
<td>.233**</td>
<td>24.0</td>
<td>Duration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at .01 level.
**Significant at .05 level.
dimensional changes which occurred in the musical excerpts with a population of adults who possessed a variety of musical experiences. Viewed in this way, the information obtained could serve as guidance for further development of a measure suitable for fourth-grade elementary children. The overall performance of this group indicated that the predominant change in each musical item was discernible. Examination of the item difficulty distribution indicated that a substantial majority of the group chose the correct answers. It was decided to use this form of the test with a group of elementary school pupils to determine whether or not the items would have appropriate discriminating power with children. Any substantial revisions could then be made.

**Trial I, Form A, June, 1969**

**Purpose.**—The purpose of this trial was to test the items and the format of the Listening Test, Form A, with elementary school pupils and to identify procedural difficulties and non-usable items.

**Subjects.**—The subjects in this trial were twenty-eight elementary band students from the Edison Middle School, Springfield, Illinois. The children were about equally divided between those who had just completed the fourth grade and those who had just completed the fifth grade.

**Materials and Procedures.**—The eighteen-item audi-taped recording used in the Exploratory Trial was used in
this trial. The tape included spoken directions and four sample items. Subjects were instructed to listen to each item, to choose the best answer, and then to listen to the excerpt again before marking the answer. After each sample item the tape was stopped to provide time for self-checking and discussion to assure understanding of the task.

Results.—The item analyses for Trial I, Form A, are presented in Tables 4, 5, and 6. The scores ranged from 6 to 16, with a mean of 10.86 and a standard deviation of 2.28.

The reliability estimate, based on the Kuder-Richardson "20" formula, was .454. This low reliability estimate was perhaps a reflection of the homogeneity of the group, length of the test, small number of subjects, and relatively limited range of scores (see Table 4).

<table>
<thead>
<tr>
<th>TABLE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIAL I; FORM A: RANGE, MEAN, STANDARD DEVIATION, AND RELIABILITY ESTIMATE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. Items</th>
<th>Range</th>
<th>Mean</th>
<th>S.D.</th>
<th>Reliability Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>6-16*</td>
<td>10.86</td>
<td>2.28**</td>
<td>.434</td>
</tr>
</tbody>
</table>

*Total score possible = 18.  
**N = 28.

Mean item difficulty for this trial was .397, indicating that it was a relatively easy test for this group of children. The majority of the items were in the "very easy" and "average" ranges. A summary of the item difficulty is
given in Table 5. The item difficulty of each item is presented in Table 6 with other data.

TABLE 5

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Mean</th>
<th>Range</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>.00- .20</td>
<td>3,4,6,9,11,17</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.21- .40</td>
<td>1,8,14,15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.41- .60</td>
<td>5,7,12,13</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.61- .80</td>
<td>2,10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.81-1.00</td>
<td>16,18</td>
<td></td>
</tr>
</tbody>
</table>

The item discrimination index for this group was based on the upper 21.43 per cent and lower 21.43 per cent of the students' mean scores. The mean item discrimination index was 35.2. Twelve items had discrimination indices of over 25 and one item was below this point. Five items had no discriminating power. Phi coefficients for eight items were significant at the .01 level of confidence and for four items at the .05 level. The point biserial r coefficients for five items were at the .01 level and two were at the .05 level. The results of the item analysis and the dimension measured are presented in Table 6.

Discussion.—In this trial, two music teachers again took the test at the same time that the students were taking it. Their responses agreed with those of the investigator. The low reliability estimate, however, and the uneven discriminating power of the items necessitated
TABLE 6
TRIAL I; FORM A: ITEM ANALYSIS

<table>
<thead>
<tr>
<th>Item</th>
<th>Diff. Level</th>
<th>Phi Coef.</th>
<th>Point Bis. r</th>
<th>Discrim. Index</th>
<th>Dimension Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.214</td>
<td>.790*</td>
<td>.426**</td>
<td>50.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>2</td>
<td>.607</td>
<td>.0</td>
<td>-.014</td>
<td>0.0</td>
<td>Pitch</td>
</tr>
<tr>
<td>3</td>
<td>.143</td>
<td>.790*</td>
<td>.601*</td>
<td>50.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>4</td>
<td>.179</td>
<td>.972*</td>
<td>.666*</td>
<td>83.3</td>
<td>Pitch</td>
</tr>
<tr>
<td>5</td>
<td>.571</td>
<td>.972*</td>
<td>.624*</td>
<td>83.3</td>
<td>Duration</td>
</tr>
<tr>
<td>6</td>
<td>.107</td>
<td>.649**</td>
<td>.232</td>
<td>33.3</td>
<td>Duration</td>
</tr>
<tr>
<td>7</td>
<td>.464</td>
<td>.0</td>
<td>.067</td>
<td>0.0</td>
<td>Loudness, Duration</td>
</tr>
<tr>
<td>8</td>
<td>.357</td>
<td>.0</td>
<td>.182</td>
<td>0.0</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>9</td>
<td>.179</td>
<td>.294</td>
<td>.298</td>
<td>16.7</td>
<td>Loudness, Duration</td>
</tr>
<tr>
<td>10</td>
<td>.679</td>
<td>.898*</td>
<td>.546*</td>
<td>66.7</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>11</td>
<td>.143</td>
<td>.790*</td>
<td>.601*</td>
<td>50.0</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>12</td>
<td>.429</td>
<td>.718*</td>
<td>.294</td>
<td>50.0</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>13</td>
<td>.536</td>
<td>.790*</td>
<td>.247</td>
<td>50.0</td>
<td>Duration</td>
</tr>
<tr>
<td>14</td>
<td>.393</td>
<td>.495</td>
<td>.238</td>
<td>33.3</td>
<td>Pitch</td>
</tr>
<tr>
<td>15</td>
<td>.250</td>
<td>.649**</td>
<td>.145</td>
<td>33.3</td>
<td>Loudness</td>
</tr>
<tr>
<td>16</td>
<td>.893</td>
<td>.649**</td>
<td>.427**</td>
<td>33.3</td>
<td>Pitch</td>
</tr>
<tr>
<td>17</td>
<td>.036</td>
<td>.0</td>
<td>-.097</td>
<td>0.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>18</td>
<td>.964</td>
<td>.0</td>
<td>.012</td>
<td>0.0</td>
<td>Duration</td>
</tr>
</tbody>
</table>

*Significant at .01 level.
**Significant at .05 level.

substantial revisions before the test could be of value as a measuring instrument for children. After careful evaluation of the testing procedures and presentation, the investigator decided that the musical excerpts were too long for the attention span of the children. In an attempt to present musical excerpts which in themselves gave a sense of aural completeness, the eighteen items took almost fifty minutes to administer. It was observed that many of the children seemed to grow restless as the test proceeded, and that often they marked their responses before hearing the complete excerpt. Therefore, a new test was developed by
retaining the usable excerpts deemed appropriate in length, shortening others, and adding new excerpts so that a usable number of items for a test of approximately forty minutes would result. It was decided to limit the excerpts to not more than forty seconds in length. The resulting items ranged from approximately five seconds to forty seconds. The added excerpts also provided a larger number of items for analysis.

**Trial II, Form B, September, 1969**

**Purpose.**—The purpose of this trial was to test the listening items with an arbitrary sample of fourth graders and to obtain data for statistical analysis.

**Subjects.**—The subjects for this trial were the fourth-grade children in the Atwater Elementary School, Circleville, Ohio. Ninety-seven students took the test. The school population was described by the teachers as generally lower-class, middle and upper-middle class, having perhaps a more middle- than lower-class composition.

**Materials and Procedures.**—The test used in this trial contained the shortened excerpts from the first listening tape, Form A, with newly added items. It is presented in Appendix B as Form B. This test contained thirty-two items. Items 1 through 13 contained a change in one dimension; items 14 through 24 included changes in two dimensions, and items 25 through 32 were paired items containing a change in one dimension. The original four
sample items were retained. The same procedure was followed as in the previous trial. Time was given after each of the sample items for self-checking and discussion. There was also a short "break" between the sections of the test to allow the children to move about, sing a song, and relax.

Results.—The item analyses for Trial II, Form B, are presented in Tables 7, 8, and 9. Scores ranged from 6 to 26, with a mean of 14.76 and a standard deviation of 4.41. The reliability estimate was .682 (see Table 7).

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Range</th>
<th>Mean</th>
<th>S.D.</th>
<th>Reliability Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>6.26*</td>
<td>14.76</td>
<td>4.41**</td>
<td>.682</td>
</tr>
</tbody>
</table>

Total score possible = 32.

**N = 97.

Mean item difficulty was .539. Item difficulty distribution was a fairly symmetrical spread. Table 8 presents the item difficulty summary, while specific item difficulty is given in Table 9.

Discrimination indices ranged from -4.1 to 76.6, with a mean of 33. Twenty-three items were above 20. Twenty-three items had phi coefficients significant at the .01 level of confidence. Point biserial r coefficients were significant at the .01 level for 21 items and at the .05
TABLE 8
TRIAL II; FORM B: SUMMARY ITEM DIFFICULTY

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Mean</th>
<th>Range</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.539</td>
<td>.21- .40</td>
<td>1, 4, 6, 9, 18, 27, 29, 32</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>.41- .60</td>
<td>2, 7, 8, 10, 14, 15, 16, 19, 21, 23, 24, 26</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>.61- .80</td>
<td>5, 12, 13, 20, 22, 25, 28, 31</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.81-1.00</td>
<td>11, 17, 30</td>
</tr>
</tbody>
</table>

level for 3 items. The data from the item analysis and the dimensions measured are presented in Table 9.

Discussion.—The analysis of the items used in this trial showed a larger number of usable items. The mean item difficulty for this form of the test was .539. This figure closely approached the 50 per cent difficulty level suggested by Downie and Heath.¹

It was thought desirable at this stage in the development of the test to compare the separate sections of the test in order to determine whether or not the three sections of the test, requiring slightly different response sets, were equal in terms of difficulty. Was it more difficult for these fourth-graders to attend to the task of detecting a two-dimensional change than a single change? Did a comparison of two closely related excerpts prove less difficult than detecting a change which occurred without a time break in the music? To provide some information, the mean

¹Downie and Heath, op. cit., p. 229.
TABLE 9

TRIAL II; FROM B: ITEM ANALYSIS

<table>
<thead>
<tr>
<th>Item</th>
<th>Diff. Level</th>
<th>Phi Coef.</th>
<th>Point Bis. r</th>
<th>Discrim. Index</th>
<th>Dimension Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.351</td>
<td>.760*</td>
<td>.397*</td>
<td>53.3</td>
<td>Loudness</td>
</tr>
<tr>
<td>2</td>
<td>.588</td>
<td>.696*</td>
<td>.335*</td>
<td>49.4</td>
<td>Pitch</td>
</tr>
<tr>
<td>3</td>
<td>.031</td>
<td>.264</td>
<td>.166</td>
<td>6.5</td>
<td>Loudness</td>
</tr>
<tr>
<td>4</td>
<td>.330</td>
<td>.637*</td>
<td>.370*</td>
<td>40.4</td>
<td>Pitch</td>
</tr>
<tr>
<td>5</td>
<td>.763</td>
<td>.707*</td>
<td>.377*</td>
<td>46.3</td>
<td>Duration</td>
</tr>
<tr>
<td>6</td>
<td>.381</td>
<td>.509*</td>
<td>.300*</td>
<td>34.1</td>
<td>Duration</td>
</tr>
<tr>
<td>7</td>
<td>.474</td>
<td>.707*</td>
<td>.338*</td>
<td>49.3</td>
<td>Loudness</td>
</tr>
<tr>
<td>8</td>
<td>.515</td>
<td>.588*</td>
<td>.282*</td>
<td>39.7</td>
<td>Duration</td>
</tr>
<tr>
<td>9</td>
<td>.330</td>
<td>.780*</td>
<td>.490*</td>
<td>54.1</td>
<td>Loudness</td>
</tr>
<tr>
<td>10</td>
<td>.536</td>
<td>.696*</td>
<td>.360*</td>
<td>49.4</td>
<td>Pitch</td>
</tr>
<tr>
<td>11</td>
<td>.866</td>
<td>.233</td>
<td>.152</td>
<td>9.5</td>
<td>Pitch</td>
</tr>
<tr>
<td>12</td>
<td>.619</td>
<td>.661</td>
<td>.456*</td>
<td>43.1</td>
<td>Loudness</td>
</tr>
<tr>
<td>13</td>
<td>.804</td>
<td>.495*</td>
<td>.321*</td>
<td>25.5</td>
<td>Pitch</td>
</tr>
<tr>
<td>14</td>
<td>.454</td>
<td>.454*</td>
<td>.380*</td>
<td>30.1</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>15</td>
<td>.485</td>
<td>.110</td>
<td>.116</td>
<td>6.8</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>16</td>
<td>.598</td>
<td>.562*</td>
<td>.340*</td>
<td>36.6</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>17</td>
<td>.835</td>
<td>.031</td>
<td>.005</td>
<td>1.4</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>18</td>
<td>.216</td>
<td>-.016</td>
<td>.091</td>
<td>-0.6</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>19</td>
<td>.546</td>
<td>.426*</td>
<td>.204**</td>
<td>27.7</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>20</td>
<td>.701</td>
<td>.613*</td>
<td>.429*</td>
<td>39.9</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>21</td>
<td>.598</td>
<td>.844*</td>
<td>.512*</td>
<td>63.1</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>22</td>
<td>.732</td>
<td>.625*</td>
<td>.276*</td>
<td>38.3</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>23</td>
<td>.495</td>
<td>.891*</td>
<td>.504*</td>
<td>67.7</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>24</td>
<td>.495</td>
<td>.930*</td>
<td>.635*</td>
<td>76.6</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>25</td>
<td>.794</td>
<td>.482*</td>
<td>.259**</td>
<td>26.3</td>
<td>Duration</td>
</tr>
<tr>
<td>26</td>
<td>.433</td>
<td>.0</td>
<td>.047</td>
<td>0.4</td>
<td>Pitch</td>
</tr>
<tr>
<td>27</td>
<td>.237</td>
<td>.575*</td>
<td>.372*</td>
<td>36.4</td>
<td>Loudness</td>
</tr>
<tr>
<td>28</td>
<td>.722</td>
<td>.294</td>
<td>.211**</td>
<td>16.6</td>
<td>Pitch</td>
</tr>
<tr>
<td>29</td>
<td>.392</td>
<td>.760*</td>
<td>.465*</td>
<td>55.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>30</td>
<td>.835</td>
<td>.078</td>
<td>.062</td>
<td>3.9</td>
<td>Duration</td>
</tr>
<tr>
<td>31</td>
<td>.794</td>
<td>-.094</td>
<td>-.054</td>
<td>-4.1</td>
<td>Pitch</td>
</tr>
<tr>
<td>32</td>
<td>.289</td>
<td>.536*</td>
<td>.348*</td>
<td>33.2</td>
<td>Duration</td>
</tr>
</tbody>
</table>

*Significant at .01 level.

**Significant at .05 level.
item difficulty for each section of the test was determined. Table 10 presents these data.

**TABLE 10**

**TRIAL II; FORM B: TEST SECTIONS MEAN DIFFICULTY**

<table>
<thead>
<tr>
<th>Type of Change</th>
<th>No. of Items</th>
<th>Mean Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-dimensional</td>
<td>13</td>
<td>.507</td>
</tr>
<tr>
<td>Two-dimensional</td>
<td>11</td>
<td>.560</td>
</tr>
<tr>
<td>Paired excerpts, one-dimensional</td>
<td>8</td>
<td>.562</td>
</tr>
</tbody>
</table>

It would appear that the separate sections were nearly equal in terms of difficulty level.

An effort was made to achieve a balance among the items measuring pitch, loudness, and duration. Though it was not the purpose of this project to present observations about the relative difficulty of the three dimensions, mean difficulty levels would perhaps provide interesting information. Table 11 presents the mean difficulty indices for the three dimensions measured.

**TABLE 11**

**TRIAL II; FORM B: DIMENSION MEAN DIFFICULTY**

<table>
<thead>
<tr>
<th>Dimension Measured</th>
<th>Number of Items</th>
<th>Mean Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch</td>
<td>8</td>
<td>.634</td>
</tr>
<tr>
<td>Loudness</td>
<td>7</td>
<td>.348</td>
</tr>
<tr>
<td>Duration</td>
<td>6</td>
<td>.596</td>
</tr>
<tr>
<td>Duration, Loudness</td>
<td>5</td>
<td>.575</td>
</tr>
<tr>
<td>Pitch, Loudness</td>
<td>3</td>
<td>.577</td>
</tr>
<tr>
<td>Pitch, Duration</td>
<td>3</td>
<td>.515</td>
</tr>
</tbody>
</table>
The second section of the test required a response describing two changes. It was impossible to determine whether a child's incorrect answer was revealing semantic confusion (e.g., did the response "lower and softer" merely indicate the child's response to the dynamic change?) or an inability to respond verbally to two changes. In the first and third sections of the test, however, pitch discrimination appeared to be the most difficult, and loudness the least difficult of the dimensions measured.

The children in this trial were very responsive. Rapport was easily established. They responded readily to the sample items, and appeared to understand the task. They were quiet and attentive during the test. They did not appear frustrated by the test; there were comments at the end of the testing period about it being "fun" and "not hard."

It was decided to revise the test by omitting the items which the item analysis had indicated non-useful, and to seek a school population which included children from a wider socioeconomic range for a final trial.

**Trial III, Form C, September, 1969**

**Purpose.**—The purpose of this trial was to test the items in the revised Listening Test with a different group of fourth-grade subjects and to obtain additional information concerning the difficulty and discriminating power of the individual items before using them in the pilot study.
Subjects.--The subjects for this trial were 114 fourth-grade pupils in the Mound St. Elementary School and the Court St. Elementary School in Circleville, Ohio. The four classrooms represented the total fourth-grade populations in these schools. These school populations were described as ranging from upper-middle-class to lower-class, with perhaps fewer upper-middle-class children than were found in the preceding trial's school population.

Materials and Procedures.--The instrument for this trial was the Listening Test designated as Form C in Appendix C. The twenty-four items on this test were those from the Form B measure which had appropriate discrimination indices and difficulty levels. The procedures used were identical to those used in Trial II.

Results.--The mean test score for this trial was 8.58, with a standard deviation of 3.50. Scores ranged from 1 to 20. The Kuder-Richardson "20" reliability estimate was .641. Table 12 presents these data.

TABLE 12
TRIAL III; FORM C: RANGE, MEAN, STANDARD DEVIATION AND RELIABILITY ESTIMATE

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Range</th>
<th>Mean</th>
<th>S.D.</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>1-20*</td>
<td>8.58</td>
<td>3.50*</td>
<td>.641</td>
</tr>
</tbody>
</table>

*Total score possible = 24.
**N = 114.
Mean item difficulty was .643. Item difficulty distribution was relatively well-spread. Item difficulty ranges are shown in Table 13. Specific item difficulty is presented in Table 16 with the complete item analysis.

**TABLE 13**  
TRIAL III; FORM C: SUMMARY ITEM DIFFICULTY

<table>
<thead>
<tr>
<th>No. of Items</th>
<th>Mean</th>
<th>Range</th>
<th>Item Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00-0.20</td>
<td>5,21</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.21-0.40</td>
<td>5,21</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.41-0.60</td>
<td>1,6,7,8,9,12,18,23,24</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.51-0.80</td>
<td>2,3,10,13,14,16,17,19</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.81-1.00</td>
<td>4,11,15,20,22</td>
<td></td>
</tr>
</tbody>
</table>

Item difficulty means for the separate sections of the test are presented in Table 14. The separate sections of the test appeared to be quite equal in terms of mean difficulty.

**TABLE 14**  
TRIAL III; FORM C: TEST SECTIONS MEAN DIFFICULTY

<table>
<thead>
<tr>
<th>Type of Change</th>
<th>No. of Items</th>
<th>Mean Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-dimensional</td>
<td>11</td>
<td>.618</td>
</tr>
<tr>
<td>Two-dimensional</td>
<td>8</td>
<td>.699</td>
</tr>
<tr>
<td>Paired excerpts, one-dimensional</td>
<td>5</td>
<td>.605</td>
</tr>
</tbody>
</table>

Item difficulty of the three dimensions measured showed pitch items to be the most difficult and duration items the least difficult. Dimension difficulty means are presented in Table 15.
TABLE 15

TRIAL III; FORM C: DIMENSION MEAN DIFFICULTY

<table>
<thead>
<tr>
<th>Dimension Measured</th>
<th>Number of Items</th>
<th>Mean Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch</td>
<td>5</td>
<td>.707</td>
</tr>
<tr>
<td>Loudness</td>
<td>6</td>
<td>.621</td>
</tr>
<tr>
<td>Duration</td>
<td>5</td>
<td>.531</td>
</tr>
<tr>
<td>Duration, Loudness</td>
<td>5</td>
<td>.659</td>
</tr>
<tr>
<td>Pitch, Loudness</td>
<td>2</td>
<td>.868</td>
</tr>
<tr>
<td>Pitch, Duration</td>
<td>1</td>
<td>.561</td>
</tr>
</tbody>
</table>

The mean item discrimination was 35.1. Seven items fell in the range below 20, seven items were in the 21-40 range, eight items were in the 41-60 range, and two in the 61-80 range. Eighteen items had phi coefficients significant at the .01 level, and two at the .05 level. Point biserial r coefficients were significant at the .01 level for seventeen items. One item was significant at the .05 level. Table 16 presents the item analysis and dimensions measured.

Discussion.—The testing conditions in one of the schools proved less than ideal in this trial. In one classroom, outside noise disturbance made it necessary to stop the tape for a short interval until it was again possible to hear the musical excerpts. This resulted in a lapse in attention by some of the children. In another classroom, a disciplinary problem between the classroom teacher and a group of children who had arrived after the bell had rung caused difficulty in establishing a suitable "testing climate." It was observed that several children
### TABLE 16

**TRIAL III; FORM C: ITEM ANALYSIS**

<table>
<thead>
<tr>
<th>Item</th>
<th>Diff. Level</th>
<th>Phi Coef.</th>
<th>Point Bis. r</th>
<th>Discrim. Index</th>
<th>Dimension Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.553</td>
<td>.836*</td>
<td>.527*</td>
<td>62.7</td>
<td>Loudness</td>
</tr>
<tr>
<td>2</td>
<td>.649</td>
<td>.309**</td>
<td>.167</td>
<td>18.1</td>
<td>Pitch</td>
</tr>
<tr>
<td>3</td>
<td>.781</td>
<td>.800*</td>
<td>.561*</td>
<td>55.4</td>
<td>Pitch</td>
</tr>
<tr>
<td>4</td>
<td>.877</td>
<td>.187</td>
<td>.083</td>
<td>7.0</td>
<td>Duration</td>
</tr>
<tr>
<td>5</td>
<td>.254</td>
<td>.495*</td>
<td>.275*</td>
<td>30.9</td>
<td>Duration</td>
</tr>
<tr>
<td>6</td>
<td>.482</td>
<td>.575*</td>
<td>.321*</td>
<td>38.6</td>
<td>Loudness</td>
</tr>
<tr>
<td>7</td>
<td>.588</td>
<td>.685*</td>
<td>.396*</td>
<td>46.6</td>
<td>Duration</td>
</tr>
<tr>
<td>8</td>
<td>.518</td>
<td>.729*</td>
<td>.463*</td>
<td>51.5</td>
<td>Loudness</td>
</tr>
<tr>
<td>9</td>
<td>.482</td>
<td>.575*</td>
<td>.341*</td>
<td>38.9</td>
<td>Pitch</td>
</tr>
<tr>
<td>10</td>
<td>.798</td>
<td>.661*</td>
<td>.367*</td>
<td>39.3</td>
<td>Loudness</td>
</tr>
<tr>
<td>11</td>
<td>.816</td>
<td>.468*</td>
<td>.264*</td>
<td>23.1</td>
<td>Pitch</td>
</tr>
<tr>
<td>12</td>
<td>.456</td>
<td>.309**</td>
<td>.217**</td>
<td>20.1</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>13</td>
<td>.728</td>
<td>.707*</td>
<td>.429*</td>
<td>46.0</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>14</td>
<td>.798</td>
<td>.718*</td>
<td>.417*</td>
<td>45.7</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>15</td>
<td>.939</td>
<td>.187</td>
<td>.125</td>
<td>7.0</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>16</td>
<td>.675</td>
<td>.729*</td>
<td>.469*</td>
<td>49.5</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>17</td>
<td>.763</td>
<td>.156</td>
<td>.120</td>
<td>8.4</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>18</td>
<td>.561</td>
<td>.780*</td>
<td>.455*</td>
<td>56.5</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>19</td>
<td>.675</td>
<td>.853*</td>
<td>.544*</td>
<td>61.9</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>20</td>
<td>.886</td>
<td>.397*</td>
<td>.162</td>
<td>16.4</td>
<td>Duration</td>
</tr>
<tr>
<td>21</td>
<td>.246</td>
<td>.495*</td>
<td>.287*</td>
<td>28.6</td>
<td>Loudness</td>
</tr>
<tr>
<td>22</td>
<td>.807</td>
<td>.218</td>
<td>.065</td>
<td>11.1</td>
<td>Pitch</td>
</tr>
<tr>
<td>23</td>
<td>.588</td>
<td>.637*</td>
<td>.351*</td>
<td>44.2</td>
<td>Loudness</td>
</tr>
<tr>
<td>24</td>
<td>.500</td>
<td>.522*</td>
<td>.296*</td>
<td>35.4</td>
<td>Duration</td>
</tr>
</tbody>
</table>

*Significant at .01 level.

**Significant at .05 level.**

In this classroom seemed to give little attention to the test. These incidents perhaps affected the performances of the children in these rooms, and thereby had some effect on the total results. The range of scores in this trial was lower, as was the mean score. In the other two classrooms, however, there seemed to be the same careful attention and interest in the task that had characterized the preceding trial's testing situation.
The music teacher of this group of children and a classroom teacher with some musical experience took the test with the students. The music teacher disagreed with the investigator's judgment on one pitch item, and the classroom teacher disagreed on three items. In view of these results, items 15, 22, and 24 would seem to be questionable. Item 15 proved to be the most difficult item on the test for the children, also. The pitch change involved a melodic sequence, one step higher and second time the motive sounded, and softer in dynamic level. Items 22 and 24 had difficulty levels of .807 and .500, respectively, for the children. Item 24 had a phi coefficient and point biserial r coefficient significant at the .01 level.

Several decisions were made at this stage in the development of the Listening Test. The reliability estimate of .641 was not as high as should be expected for a standardized test, but met the standard for group measurement that Nunnally has suggested.¹ Factors which perhaps influenced the reliability of this measure included the amount of guessing the children did, and the quality of attention which they gave to the test. Though a relatively wide socioeconomic range was sought in this sample, it is perhaps realistic to assume that extremes at either end of a socioeconomic continuum were not included in this small-

¹Nunnally, op. cit., p. 226.
town school population. This factor perhaps tended to make the sample more homogeneous than a sample drawn from a wider geographic area with more widely varying socioeconomic levels.

Nunnally has stated that, in terms of item discrimination, it is sensible to construct tests primarily considering the correlation of items with total scores.\(^1\) In this form of the test, all except six items were found to be significant at the .01 level of confidence, and one of these six items was significant at the .05 level when the point biserial \(r\) was computed. With the exception of one item, all items which were not at the .01 level in this trial were at that level in the previous trial. For the important purpose of maintaining balance among the dimensions measured, however, the investigator decided to use the test in its latest form for the pilot study. It appeared to be acceptable for the purpose for which it was developed.

Validation Trial, Form C

**Purpose.**—The purpose of this trial was to ascertain whether or not the items selected for the final Listening Test would elicit correct responses indicating the musical change.

**Subjects.**—The subjects for this trial were 51 elementary education majors enrolled in a course dealing with

music methods and materials. Thirty-three of these students indicated that they had had some previous training in band, orchestra, chorus, or piano beyond the grade school level.

Materials and Procedures.—The Listening Test, Form C, was used in this trial. The procedures were the same as for the previous trials.

Results.—Table 17 presents the percentage of students who answered the item correctly and the musical dimension measured. The column heading, "N = 51" indicates the percentage of correct answers from the whole group and "N = 33" indicates the percentage of correct answers from the group who had had some musical training. Four students chose the correct responses for all items, and ten students missed one item only.

Discussion.—A high proportion of correct answers was achieved by the group. Though several weak items were identified, it was decided that the test as a whole contained items in which the musical changes were discernible. It was decided to use the Listening Test, Form C, for the pilot study.

Summary

The final Listening Test developed for use in a pilot study seemed to have the following characteristics thought to be important to the purposes of the main study:

1. The test appeared to be appropriate in length. It could be administered with ease in a 45-minute class
### TABLE 17

PERCENTAGE OF CORRECT ANSWERS LISTENING TEST, FORM C VALIDATION TRIAL

<table>
<thead>
<tr>
<th>Item Number</th>
<th>N=51 Per Cent</th>
<th>N=33 Per Cent</th>
<th>Dimension Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.1</td>
<td>93.9</td>
<td>Loudness</td>
</tr>
<tr>
<td>2</td>
<td>82.4</td>
<td>90.9</td>
<td>Pitch</td>
</tr>
<tr>
<td>3</td>
<td>91.4</td>
<td>97.0</td>
<td>Pitch</td>
</tr>
<tr>
<td>4</td>
<td>86.3</td>
<td>93.9</td>
<td>Duration</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>100.0</td>
<td>Duration</td>
</tr>
<tr>
<td>6</td>
<td>82.4</td>
<td>84.8</td>
<td>Loudness</td>
</tr>
<tr>
<td>7</td>
<td>98.0</td>
<td>100.0</td>
<td>Duration</td>
</tr>
<tr>
<td>8</td>
<td>98.0</td>
<td>97.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>9</td>
<td>98.0</td>
<td>97.0</td>
<td>Pitch</td>
</tr>
<tr>
<td>10</td>
<td>94.1</td>
<td>97.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>11</td>
<td>53.0</td>
<td>63.6</td>
<td>Pitch</td>
</tr>
<tr>
<td>12</td>
<td>86.3</td>
<td>93.9</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>13</td>
<td>94.1</td>
<td>100.0</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>14</td>
<td>82.4</td>
<td>87.9</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>15</td>
<td>53.0</td>
<td>51.6</td>
<td>Pitch, Loudness</td>
</tr>
<tr>
<td>16</td>
<td>92.2</td>
<td>93.9</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>17</td>
<td>80.0</td>
<td>84.8</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>18</td>
<td>96.1</td>
<td>93.9</td>
<td>Pitch, Duration</td>
</tr>
<tr>
<td>19</td>
<td>96.1</td>
<td>93.9</td>
<td>Duration, Loudness</td>
</tr>
<tr>
<td>20</td>
<td>86.3</td>
<td>93.9</td>
<td>Duration</td>
</tr>
<tr>
<td>21</td>
<td>94.1</td>
<td>93.9</td>
<td>Loudness</td>
</tr>
<tr>
<td>22</td>
<td>64.7</td>
<td>75.8</td>
<td>Pitch</td>
</tr>
<tr>
<td>23</td>
<td>90.2</td>
<td>90.0</td>
<td>Loudness</td>
</tr>
<tr>
<td>24</td>
<td>94.1</td>
<td>93.9</td>
<td>Duration</td>
</tr>
</tbody>
</table>

period, and it did not seem to exceed the maximum attention span of fourth-grade children.

2. The format of the response sheet appeared to be appropriate for fourth-grade children. They had no trouble in following the directions, or in maintaining the pace of the test.

3. The musical items were drawn from orchestral compositions which are cited as desirable listening experiences for children in the elementary grades. Music
educators are guided in current music texts in presenting basic musical concepts by using this literature.

4. The test items provided a relatively balanced measure of the musical dimensions identified by Andrews and Diehl as important structural elements of music—pitch, duration, and loudness.

5. The items had shown acceptable mean discrimination indices and difficulty levels with arbitrarily selected fourth-grade populations.

6. The test was considered to have demonstrated sufficient content validity to make it acceptable for the purposes of the pilot study.

7. The test was considered to have a reliability estimate which, though modest, would be acceptable in a group measure of the variable being investigated in the pilot study.

Therefore, the decision was made to use the Listening Test, Form C, in the pilot study. The limitations of the measure, however, were recognized and should be considered before any further use be made of the test. Some of these limitations are stated in Chapter V under Further Research.
CHAPTER IV

A PILOT STUDY OF TWO SCHOOLS

Following the development of a satisfactory Listening Test described in Chapter III, the procedures of this study included (1) the use of the test in a pilot study of two socioeconomically contrasting school populations, and (2) the statistical treatment of the data derived from testing the two student populations with the data obtained concerning the students' backgrounds and characteristics. These procedures will be discussed in detail in this chapter.

Procedures

The pilot study of two schools involved (1) the selection of two school populations of contrasting socio-economic backgrounds, (2) the administration of the Listening Test, Form C, in the two schools, (3) the compilation of data concerning the children tested, (4) the statistical treatment of the data obtained, and (5) the presentation of the results. These activities will be described in the following sections.
Selection of Contrasting Fourth-Grade Populations

With the cooperation and aid of the vocal music supervisor for the city of Columbus, Ohio, two schools in that city were selected for the sample populations. Columbus, Ohio, in 1970 was a city with a population of approximately 580,000, in a metropolitan area of about one million. It is situated in the central portion of Ohio.

The first school selected for this study, which shall be described as School A, was located in a predominantly middle-class section of the city. The neighborhood surrounding the school consisted of modest, well-kept homes and many apartment buildings and duplexes. Some were recently built; all were well-maintained. Parents of the children attending School A were professional, semi-professional, and "blue-collar" workers. The neighborhood's proximity to a university attracted graduate students with children to the area; the school also served the children of the university students living in the university housing complex. The school population was comprised of both white and negro students, though the white population was substantially larger. There were also a few Oriental children, and some Asian children whose parents were foreign graduate students at the university.

The second school selected, which shall be designated as School B, was located in that part of the city which could be termed the inner-city area. It was
designated by the Columbus City School District as eligible for Title I services on the basis of the following factors:

(1) Pupil Absence. The system-wide absence rate was 5 per cent. School B's pupil absence rate was 7 per cent. The highest absence rate of the 48 elementary schools eligible for Title I services was 11 per cent, the lowest 3 per cent.

(2) Pupil Mobility. The mobility rate of School B was 40 per cent. The system-wide mobility rate was 15 per cent.

(3) Pupils Above Age in Grade Level. In School B, 12 per cent of the pupils were two years or more older than the youngest legal entry age for that grade level.

(4) Proportion of White Pupils. School B's population was 70 per cent white.

(5) Staff Turnover Rate. School B's staff turnover rate was 27 per cent. The system-wide staff turnover rate was 23 per cent.

School B was labelled a "high priority" school in terms of meeting the needs of disadvantaged children. In a rank order of Title I eligible schools by percentage of pupils one year or more below grade level on a reading subtest, from most (1) to least (48), School B was ranked in third place.
The information given about School B was obtained from a report entitled "Language Development and Mathematics Improvement," issued by the Columbus City School District, Division of Instruction, August, 1969.

As School A was not so designated, and was not termed a "high-priority" school, the two schools described were assumed to have represented sufficient contrast in socioeconomic level to provide useful data for the present study.

**Administration of the Listening Test, Form C**

The administration of the Listening Test in the two schools was conducted in October and November, 1969. In School A, the testing took place in the classrooms during two mornings. The procedures followed were the same as in the trials described in Chapter III. After the task was explained, some songs were sung, and a short discussion of the terms listed on the response sheet followed. The children were attentive and responsive; their "attitudinal set" seemed good. This atmosphere prevailed throughout the testing period. During the playing of the musical items, some children moved quietly in response to the music. In the time interval designated for marking the response sheet, quiet concentration seemed to characterize the group. Several children expressed satisfaction at the end of the testing period. Comments ranged from "It was easy" to
"It was fun." No interruptions occurred during the testing period.

The testing in School B was done in the classrooms during three mornings. Identical procedures were followed. After a song was sung, a discussion of the terms listed on the response sheet followed. The terms were put on the chalk-board and discussed so that the children should experience as little difficulty as possible with the reading of the terms. It became evident in one group that confusion existed as to the meanings of the words when applied to musical events. During the discussion of the words, one child volunteered that "lower" and "softer" had the same meaning. As this problem arose without warning, a short discussion followed about the differences in meaning. It could not be assumed, however, that one short discussion would change habits of thinking that had presumably existed for some time.

These children were also responsive and friendly. During the playing of the test items, several children were observed moving to the music. The children, however, were not as quiet during the testing period as the children in School A. In two classrooms, the time interval designated for marking the response sheet was characterized by low noise level. It seemed that the children took less time to mark their answers. No interruptions occurred, and when asked whether the test had seemed difficult or easy,
several children responded that it was "fun" and that it was "not hard." No frustration with the task was voiced or revealed in any way to the investigator.

Compilation of Data

With the cooperation of the principals in the two schools involved, scores from two standardized tests were obtained for the fourth-grade students who had participated in the study. The tests, the California Short-Form Test of Mental Maturity, Level 2, Form S, for Grade 4, and the California Reading Test for Grade 4, were administered in October, 1969, as part of the annual city-wide testing program. They were given approximately two weeks preceding the dates when the Listening Test, Form C, was administered.

The California Short-Form Test of Mental Maturity was described as measuring the rate and scope of mental development in terms of four statistically-derived factors: Logical Reasoning, Numerical Reasoning, Verbal Concepts, and Memory. The derived scores were grouped to yield a language and a non-language IQ, as well as a total IQ. The total IQ scores were converted to IQ scores designed to provide a constant mean of 100 and a standard deviation of 16 points for all age levels. These IQ scores were in turn used to derive a Mental Age score for each student.

The California Reading Test contained two sub-tests, Vocabulary Comprehension and Reading Comprehension. The
vocabulary test was a 40-item test, with a standardized mean of 19.25 and a standard deviation of 8.76. The reading comprehension test was a 45-item test with a standardized mean of 21.67 and a standard deviation of 9.76. These two scores were recorded separately.

In addition to these scores, the chronological age for each student was obtained. For each student, therefore, five data were compiled--Listening Test score, chronological age, mental age, vocabulary comprehension score, and reading comprehension score. These comprised the data for the pilot study.

Treatment of Data

The first purpose of this study was to determine whether or not differences existed between lower-class and middle-class children's identification of musical concepts as measured by the Listening Test, Form C. In order to determine what differences, if any, existed, the following statistical procedures were used:

(1) Ranges, means and standard deviations for each group's performance on the Listening Test, Form C, were determined, and a t test was computed to determine the significance of difference in the mean scores of the two groups.

(2) Reliability estimates for the Listening Test in each school were determined for descriptive comparison.
(3) Item difficulty means for the total test, for each section of the test, and for each musical dimension measured by the test were computed for descriptive comparison.

(4) t tests were computed for determining significant differences in the two groups' performances on individual items of the Listening Test.

To determine whether or not significant differences existed in the two groups in other areas for which data were obtained, t tests were computed using mean scores and standard deviations from the following:

(1) Chronological Age scores in School A and School B.

(2) Mental Age scores in School A and School B.

(3) Vocabulary Comprehension scores in School A and School B.

(4) Reading Comprehension scores in School A and School B.

The second purpose of this study was to examine the relationships between the children's identification of musical concepts as measured by the Listening Test, Form C, and their chronological ages, mental ages, vocabulary comprehension, and reading comprehension as measured by scores obtained for each of these variables. The data from each school were treated separately, and the following statistical procedures were used:
(1) Correlation coefficients for all possible pairs of variables within each group were derived.

(2) Multiple regression analysis using correlation data was performed.

(3) Multiple correlation coefficients were determined to show each variable's contribution to the Listening Test scores.

(4) Partial correlation coefficients were computed to show further relationships between the variables and the Listening Test scores.

(5) F values were derived and tested to determine the significance of relationship of the variables to the Listening Test scores.

The item analyses and multiple correlations were computed by the Center for Evaluation and Measurement of the Ohio State University. The complete item analysis program is given in Appendix D; the BMD02R correlation program is presented in Appendix E. The t tests were computed by the investigator.

Results

Listening Test Results

Means, standard deviations, ranges, and reliability estimates.---In order to determine what differences existed in the identification of musical concepts as measured by the Listening Test, Form C, between School A and School B, mean scores, standard deviations, ranges, and reliability
estimates for the test scores in each school were determined. A t test was then computed to determine significant differences between the mean scores. The following formula was used:

\[ t = \frac{\bar{x}_1 - \bar{x}_2}{S_{D\bar{x}}} \]

where \( S_{D\bar{x}} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} \)

\( S_{D\bar{x}} \) = the standard error of the difference between two means for uncorrelated data.

\( s_{\bar{x}_1}, s_{\bar{x}_2} \) = the standard errors of the two sample means.\(^1\)

Table 18 presents these data.

**Table 18**

<table>
<thead>
<tr>
<th>School</th>
<th>Mean</th>
<th>S.D.</th>
<th>Range</th>
<th>Rel.</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>10.77*</td>
<td>4.18**</td>
<td>4-20</td>
<td>.742</td>
<td>5.587</td>
<td>.001</td>
</tr>
<tr>
<td>School B</td>
<td>6.44*</td>
<td>2.08***</td>
<td>3-13</td>
<td>.134</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^\text{Total Score possible} = 24.\)

\(^\text{N}=35.\)

\(^\text{N}=55.\)

An examination of Table 18 shows that differences between the mean scores in School A and School B did exist. The resulting t value of 5.587 was significant beyond the .001 level of confidence.

\(^\text{1}^\text{Downie and Heath, op. cit., pp. 132, 141.}\)
The range of scores in School A was from 4 to 20, and in School B, from 3 to 13. The Kuder-Richardson "20" reliability estimate was .742 for the Listening Test in School A. This estimate was higher than that obtained in the final trial, Trial III, described in Chapter III. In School B, however, the reliability estimate .134 would seem to indicate that the test was not an appropriate instrument for this group of children, and subsequent statistical interpretation was made recognizing this limitation.

**Item Difficulty.**—An examination of the item difficulty indices for the two schools revealed the greater degree of difficulty of the test for the children in School B. Table 19 presents the mean difficulty for the total test in each school.

<p>| Table 19 |
|------------------|------------------|
| <strong>LISTENING TEST; FORM C: MEAN DIFFICULTY</strong> |
| <strong>SCHOOL A AND SCHOOL B</strong> |</p>
<table>
<thead>
<tr>
<th>School</th>
<th>Mean Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>.551</td>
</tr>
<tr>
<td>School B</td>
<td>.732</td>
</tr>
</tbody>
</table>

Mean difficulty for School A was relatively close to the 50 per cent mean difficulty level recommended by Downie and Heath.\(^1\) Item difficulty distribution for School A was relatively well spread, with the majority of the items in

\(^{1}\text{Ibid.}, p. 229.\)
the "average" range. In School B, the majority of the items were in the range described as very difficult. Table 20 presents the item difficulty distributions in School A and School B.

**TABLE 20**

LISTENING TEST; FORM C: SUMMARY ITEM DIFFICULTY DISTRIBUTION SCHOOL A AND SCHOOL B

<table>
<thead>
<tr>
<th>Difficulty Range</th>
<th>No. of items School A</th>
<th>No. of items School B</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00-.20</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>.21-.40</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>.41-.60</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>.61-.80</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>.81-1.00</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

Mean item difficulty indices for the separate sections of the test gave further information about the relative difficulty. The first section of the test required identification of a one-dimensional change, the second section a two-dimensional change, and the third section a one-dimensional change in the second excerpt in a pair of musical excerpts. The second section of the test was the most difficult for all the children, and the third section the least difficult. Table 21 presents these data.
TABLE 21
LISTENING TEST; FORM C: TEST SECTION MEAN DIFFICULTY
SCHOOL A AND SCHOOL B

<table>
<thead>
<tr>
<th>Type of Change</th>
<th>School A</th>
<th>School B</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-dimensional</td>
<td>.500</td>
<td>.712</td>
</tr>
<tr>
<td>Two-dimensional</td>
<td>.657</td>
<td>.836</td>
</tr>
<tr>
<td>Paired, one-dimensional</td>
<td>.497</td>
<td>.639</td>
</tr>
</tbody>
</table>

Table 22 shows the mean item difficulty for the musical dimensions measured. Items involving pitch were the most difficult, and those involving loudness the least difficult.

TABLE 22
LISTEN TEST; FORM C: DIMENSION MEAN DIFFICULTY
SCHOOL A AND SCHOOL B

<table>
<thead>
<tr>
<th>Dimension Measured</th>
<th>School A</th>
<th>School B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loudness</td>
<td>.410</td>
<td>.612</td>
</tr>
<tr>
<td>Pitch</td>
<td>.606</td>
<td>.781</td>
</tr>
<tr>
<td>Duration</td>
<td>.457</td>
<td>.655</td>
</tr>
<tr>
<td>Duration, Loudness</td>
<td>.628</td>
<td>.793</td>
</tr>
<tr>
<td>Pitch, Loudness</td>
<td>.815</td>
<td>.937</td>
</tr>
<tr>
<td>Pitch, Duration</td>
<td>.855</td>
<td>.855</td>
</tr>
</tbody>
</table>

To determine the significance of difference between the difficulty indices of individual items in School A and School B, t tests for testing differences between proportions were performed, using the following formula: 1

1Ibid., p. 149.
\[ t = \frac{p_1 - p_2}{s_{dp}} \]

where \( s_{dp} = \sqrt{pq \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \)

Table 23 presents these results. A \( t \) value of 3.42 was significant at the .001 level of confidence, 2.64 at the .01 level, and 1.99 at the .05 level. Five items had \( t \) ratios significant at the .01 level, and seven at the .05 level. Ten items were not significant (see Table 23).

**TABLE 23**

LISTENING TEST; FORM C: DIFFERENCES BETWEEN PROPORTIONS OF CORRECT ANSWERS IN SCHOOL A AND SCHOOL B

<table>
<thead>
<tr>
<th>Item No.</th>
<th>Percentage of Correct Answers</th>
<th>School A*</th>
<th>School B**</th>
<th>( t )</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62.9</td>
<td>25.5</td>
<td>3.52</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>48.6</td>
<td>43.6</td>
<td>.46</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>45.7</td>
<td>18.2</td>
<td>2.79</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>28.6</td>
<td>3.6</td>
<td>3.40</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>77.1</td>
<td>49.1</td>
<td>2.63</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>48.6</td>
<td>40.0</td>
<td>.80</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>51.4</td>
<td>50.9</td>
<td>.95</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>71.4</td>
<td>45.5</td>
<td>2.40</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>51.4</td>
<td>20.0</td>
<td>3.09</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>40.0</td>
<td>12.7</td>
<td>2.98</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>25.7</td>
<td>7.3</td>
<td>1.10</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>48.6</td>
<td>47.3</td>
<td>.12</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>28.6</td>
<td>10.9</td>
<td>2.09</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>20.0</td>
<td>3.6</td>
<td>2.53</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>17.1</td>
<td>9.1</td>
<td>1.13</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>34.3</td>
<td>18.2</td>
<td>1.73</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>25.7</td>
<td>7.3</td>
<td>2.41</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>51.4</td>
<td>14.5</td>
<td>3.75</td>
<td>.001</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>48.6</td>
<td>20.0</td>
<td>2.85</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20.0</td>
<td>3.6</td>
<td>2.53</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>91.4</td>
<td>70.9</td>
<td>2.32</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>25.7</td>
<td>18.2</td>
<td>.74</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>40.0</td>
<td>38.2</td>
<td>.17</td>
<td>N.S.</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>74.3</td>
<td>65.5</td>
<td>.89</td>
<td>N.S.</td>
<td></td>
</tr>
</tbody>
</table>

*\( N=35 \).*

**\( N=55 \).**
Means, Standard Deviations and t values for Other Variables. — To provide further information about possible differences between the fourth-grade populations in School A and School B, $t$ ratios were computed to determine whether or not significant differences existed in the other areas for which data were obtained. Table 24 presents the means, standard deviations, $t$ ratio values and level of significance for the scores obtained in School A and School B for chronological age (in months), mental age (in months), vocabulary comprehension and reading comprehension. It will be noted that the number of subjects in School B is 51. Because of incomplete data, scores for four children in School B could not be used. In School A, complete data were obtained for all subjects.

**TABLE 24**

**MEANS, STANDARD DEVIATIONS, $t$ VALUES, LEVEL OF SIGNIFICANCE, OTHER VARIABLES: SCHOOL A AND SCHOOL B**

<table>
<thead>
<tr>
<th>Variable</th>
<th>School A Mean</th>
<th>S.D.</th>
<th>School B Mean</th>
<th>S.D.</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. A.</td>
<td>113.43</td>
<td>6.22</td>
<td>117.29</td>
<td>7.33</td>
<td>2.26</td>
<td>.05</td>
</tr>
<tr>
<td>M. A.</td>
<td>121.74</td>
<td>14.43</td>
<td>111.75</td>
<td>10.96</td>
<td>3.42</td>
<td>.001</td>
</tr>
<tr>
<td>Vocab.</td>
<td>21.03</td>
<td>7.98</td>
<td>14.75</td>
<td>6.65</td>
<td>3.79</td>
<td>.001</td>
</tr>
<tr>
<td>Read.</td>
<td>26.29</td>
<td>9.26</td>
<td>15.76</td>
<td>6.39</td>
<td>5.76</td>
<td>.001</td>
</tr>
</tbody>
</table>

$^*$N=35.

$^{**}$N=51.

The $t$ values indicate significant differences in School A and School B for all variables. For 84 degrees of freedom, the critical ratio for the .05 level of confidence
is 1.991, for the .01 level, 2.618, and for the .001 level, 3.427. The differences in mean scores for mental age, vocabulary comprehension and reading comprehension were significant at the .001 level of confidence, and for chronological age at the .05 level.

Multiple Regression Analysis Results

Intercorrelations: School A.—Multiple regression analysis as computed in the BMD02R program (see Appendix E) was utilized to examine the proportion of the criterion variance (Listening Test, Form C) accounted for by the complete battery of variables and by portions of the battery.

For each set of scores (School A and School B), a correlation matrix was derived. The correlation coefficients for School A are shown in Table 25.

TABLE 25
RELATIONSHIPS OF VARIOUS FACTORS TO LISTENING TEST SCORES IN SCHOOL A—(COEFFICIENTS OF CORRELATION)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chronological Age</td>
<td>.009</td>
<td>-1.138</td>
<td>-.057</td>
<td>-.012</td>
<td></td>
</tr>
<tr>
<td>2. Mental Age</td>
<td>.748</td>
<td>.771</td>
<td>.646</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocab. Comprehension</td>
<td>.811</td>
<td>.558</td>
<td>.600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Reading Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y. Listening Test, Form C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These coefficients were used in a sequence of multiple linear regression equations in a stepwise manner to ascertain the relative predictive value of each variable to
musical concept identification as measured by the Listening Test, Form C. At each step one variable was added to the regression equation. The added variable was the one which made the greatest reduction in the error sum of squares and which had the highest F value. The F values were tested for significance. In summary, this program enabled some insight into how each variable or combination of variables could be viewed as predictors of musical concept identification as measured by the Listening Test, Form C.

An examination of Table 25 indicates that variable 2, or mental age, had the highest correlation with the dependent variable, the Listening Test, in School A. Restated, it was found that in School A, mental age was the best predictor of musical concept identification as measured by the Listening Test. The F value derived from the analysis of variance was 23.609, which for 1-33 degrees of freedom was significant beyond the .01 level of confidence.

Partial correlation coefficients resulting when mental age was held constant are shown in Table 26. Variable 4, reading comprehension, had the highest partial correlation coefficient, but when its t value was determined, it was found to be significant only at the .10 level of confidence. It can thus be stated that when mental age is held constant, no other measured factor added significantly to musical concept identification achievement in School A.
TABLE 26

RELATIONSHIP OF OTHER FACTORS TO LISTENING TEST SCORES WITH MENTAL AGE SCORES HELD CONSTANT
SCHOOL A—(PARTIAL CORRELATION COEFFICIENTS)

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological Age</td>
<td>-.02345</td>
</tr>
<tr>
<td>Vocabulary Comprehension</td>
<td>.14874</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.20938</td>
</tr>
</tbody>
</table>

A summary of the multiple regression analysis for School A is presented in Table 27. Cumulative multiple correlation coefficients for the variables' relationship with the Listening Test scores are shown in the first column. The cumulative percentages for the amount of variance of the Listening Test scores, accounted for by the addition of each variable to the regression equation, are given in the second column. The amount of variance added by each variable is shown in the third column. The F values used for determining the significance of the coefficient after addition of each variable, and the resulting level of significance are shown in columns 4 and 5.

When mental age scores were held constant, no other measured factor contributed significantly to musical concept identification as measured by the Listening Test, Form C. Mental age accounted for 41.7 per cent of the total variance of the Listening Test scores. The complete battery of variables accounted for 44.4 per cent of the
TABLE 27

SUMMARY: RELATIONSHIP BETWEEN LISTENING TEST SCORES AND VARIOUS PREDICTORS: SCHOOL A

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Age</td>
<td>.6458</td>
<td>41.7</td>
<td>.4170</td>
<td>23.6086</td>
<td>.01</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.6653</td>
<td>44.3</td>
<td>.0256</td>
<td>1.4673</td>
<td>N.S.</td>
</tr>
<tr>
<td>Vocabulary Comprehension</td>
<td>.6660</td>
<td>44.4</td>
<td>.0009</td>
<td>.0502</td>
<td>N.S.</td>
</tr>
<tr>
<td>Chronological Age</td>
<td>.6660</td>
<td>44.4</td>
<td>.0000</td>
<td>.0009</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

Column 1= Multiple correlation coefficients.
Column 2= Contribution of variables to total variance (percentage).
Column 3= Amount of variance added by each variable.
Column 4= F value.
Column 5= Level of significance.

total variance of the Listening Test scores. The remaining variance, or that for which no account was made, was due to either random variation or the failure to include the independent variables that make these contributions.

Intercorrelations: School B.—Table 28 presents the intercorrelation matrix derived from the complete battery of scores in School B.

TABLE 28

RELATIONSHIP OF VARIOUS FACTORS TO LISTENING TEST SCORES IN SCHOOL B (COEFFICIENTS OF CORRELATION)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chronological Age</td>
<td>-.120</td>
<td>-.306</td>
<td>-.264</td>
<td>-.272</td>
<td></td>
</tr>
<tr>
<td>2. Mental Age</td>
<td>.626</td>
<td>.601</td>
<td>-.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocab. Comprehension</td>
<td>.800</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Reading Comprehension</td>
<td>.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Listening Test, Form C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These coefficients were used in the stepwise regression analysis described previously to ascertain the relative importance of each variable to musical concept identification as measured by the Listening Test, Form C. An examination of Table 28 reveals that the highest correlation between Listening Test scores and other variables was an inverse relationship with chronological age. Inverse relationships were also found between chronological age and all other variables, indicating that in School B the younger students, or those who had not repeated grades, achieved the highest scores on the Listening Test. The F value derived from the analysis of variance for this coefficient was 3.912, which for 1-49 degrees of freedom was not significant.

Partial correlation coefficients resulting when chronological age scores were held constant showed no significant relationship between any of the variables and Listening Test scores (see Table 29).

### TABLE 29

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Age</td>
<td>-.04377</td>
</tr>
<tr>
<td>Vocabulary Comprehension</td>
<td>-.08693</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>-.05153</td>
</tr>
</tbody>
</table>
Table 30 presents the summary of the multiple regression analysis for School B. Cumulative multiple correlation coefficients for the variables with Listening Test scores are shown in the first column. The second column shows the accumulative contribution of each variable to the total variance of the Listening Test scores. The individual contribution of each variable to the total variance is shown in the third column. F values, and the level of significance for those values as predictors of musical concept identification achievement are shown in the fourth and fifth columns.

**TABLE 30**

**SUMMARY: RELATIONSHIP BETWEEN LISTENING TEST SCORES AND VARIOUS PREDICTORS: SCHOOL B**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronological Age</td>
<td>.2719</td>
<td>7.39</td>
<td>.0739</td>
<td>3.9121</td>
<td>N.S.</td>
</tr>
<tr>
<td>Vocabulary Comprehension</td>
<td>.2845</td>
<td>8.09</td>
<td>.0070</td>
<td>.3654</td>
<td>N.S.</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>.2856</td>
<td>8.16</td>
<td>.0007</td>
<td>.0337</td>
<td>N.S.</td>
</tr>
<tr>
<td>Mental Age</td>
<td>.2857</td>
<td>8.16</td>
<td>.0001</td>
<td>.0028</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

Column 1 = Multiple correlation coefficients.  
Column 2 = Contribution of variables to total variance (percentage).  
Column 3 = Amount of variance added by each variable.  
Column 4 = F value.  
Column 5 = Level of significance.

The highest relationship between musical concept identification as measured by the Listening Test, Form C, and any of the other measured factors was an inverse one—that between chronological age and Listening Test scores. The F value derived from the analysis of variance was not
significant, however. The complete battery of variables accounted for 8.16 per cent of the total variance of the Listening Test scores. The remaining variance, or that for which no account was made, was due to either random variation or the failure to include the independent variables that make these contributions.

A summary of the findings presented in this chapter and conclusions based on the data follow in Chapter V.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

This investigation was an attempt to determine the role of environment in children's identification of selected musical concepts through aural experience with standard orchestral literature. The study used a Listening Test constructed by the investigator for use in a pilot study of two fourth-grade populations of contrasting socioeconomic backgrounds.

The Purpose of the Study

Main Purpose.—The main purpose of this study was (1) to determine whether or not children from contrasting socioeconomic backgrounds differed in musical concept development, and (2) to determine whether or not significant relationships among chronological age, mental age, vocabulary comprehension, reading comprehension, and musical concept identification existed in each group.

The study investigated the following:

(1) Is there a significant difference between lower-class and middle-class children's identification
of musical concepts as measured by the Listening Test developed for this study?

(2) Are there significant relationships among the mental ages, chronological ages, vocabulary comprehension and reading comprehension of middle-class children and their ability to identify musical concepts within the frame of reference of standard orchestral literature?

(3) Are there significant relationships among the mental ages, chronological ages, vocabulary comprehension and reading comprehension of lower-class children and their ability to identify musical concepts within the frame of reference of standard orchestral literature?

Sub-Purpose.—The sub-purpose of this study was to design and construct a Listening Test which would effectively measure differences or lack of differences in the ability to identify musical concepts by children from contrasting socioeconomic environments.

Need for the Study

A considerable amount of current and recent literature in the professional journals and texts indicated a need for a music curriculum based on a conceptual approach. Such an approach has been emphasized as a means of developing basic musical understandings which would make music more meaningful to all children. Yet despite the abundance
of literature dealing with the need for teaching for conceptual development in music, a review of related research revealed few investigations measuring such development and little research dealing directly with musical concept development.

The role of environment in conceptual development has been explored quite extensively in most school subject areas, yet there appeared to be little evidence of such exploration in music education. Recent professional journals have devoted much attention to the problems in music education related to environment, but research dealing with these problems is lacking.

Methods and Procedures

The Listening Test used in the pilot study was designed for this investigation. The instrument consisted of musical excerpts taken from standard orchestral literature suggested as appropriate for children's "listening experiences" in current elementary music texts. The excerpts demonstrated musical concepts of pitch, duration, and loudness. The test was submitted to four trial administrations and item analyses for refinement before it was used in a pilot study of two fourth-grade populations of contrasting socioeconomic level.

Following the development of the test, it was administered to a fourth-grade middle-class sample population and a fourth-grade lower-class sample population.
Information concerning mental age, chronological age, vocabulary comprehension, and reading comprehension was obtained for each student tested.

The data were treated in several ways. Item analyses of the children's performances on the test provided information about the test's relative reliability and difficulty in each group. Using $t$ tests, statistical comparisons were made of the following variables: (1) scores of the Listening Test, (2) mental ages, (3) chronological ages, (4) vocabulary comprehension scores, and (5) reading comprehension scores for each group. Multiple regression analysis provided information about the relationships between these variables within each group.

Summary of the Findings

Main Purpose. — For investigating the main purpose, three null hypotheses were developed. The findings are summarized below.

1. There is no significant difference between lower-class and middle-class fourth grade children's identification of musical concepts found in standard orchestral literature.

The first null hypothesis was rejected. A $t$ test showed that the middle-class fourth grade children scored significantly higher on the Listening Test. Values of $t$ resulting from testing the differences between proportions of correct responses on individual items were significant.
beyond the .05 level of confidence for fourteen of the twenty-four items, indicating that the test was more difficult for the lower-class children than for the middle-class children.

2. There are no significant relationships among mental age, chronological age, vocabulary comprehension, and reading comprehension of middle-class fourth-grade children and their ability to identify musical concepts found in standard orchestral literature.

Multiple regression analysis revealed a significant relationship between mental age and musical concept identification. Though the other variables, with the exception of chronological age, contributed to musical concept identification, significant relationships were not found between musical concept identification and the other variables when mental age was held constant.

3. There are no significant relationships among mental age, chronological age, vocabulary comprehension, and reading comprehension of lower-class fourth-grade children and their ability to identify musical concepts found in standard orchestral literature.

The null hypothesis was accepted. No significant relationships were found between musical concept identification and any of the variables.
Sub-Purpose.--To investigate the sub-purpose of the study, the attempt was made to construct and develop a test consisting of recorded orchestral musical excerpts which would effectively measure differences or lack of differences in identification of musical concepts by children from contrasting socioeconomic environments. Through the use of appropriate test construction methods, such a Listening Test was refined for use in the pilot study. Though the test proved acceptable in terms of reliability and difficulty for middle-class children, it did not prove so for lower-class children.

Conclusions and Recommendations

The following conclusions are formulated on the basis of the findings of this study and in consideration of the limitations of the investigation.

(1) The results of the Listening Test indicate that middle-class children do appear to be more efficient in "labeling" musical events than do their lower-class peers. Whether this ability results from a higher degree of fluency with the language, more experience with abstract tasks, or more aural experience with the type of music used in the Listening Test was not determined. One may speculate, however, that the research findings dealing with the auditory and motivational aspects of lower-class children in
other subject areas may be applicable to music listening experiences, also.

(2) The multiple correlation analysis of the lower-class group's performance on the Listening Test yielded no significant relationships among musical concept identification and mental age, chronological age, vocabulary comprehension, or reading comprehension. Considering the research findings reviewed in Chapter II, one is led to conjecture whether or not the lack of aural experience with the type of music presented in the Listening Test could account for the evident lack of comprehension of the task. Such conjecture might lead to the following conclusions:

(a) Materials and methods used in school music experiences should be chosen in accordance with the children's previous experiences with music.

(b) Exposure to unfamiliar types of music should precede any attempt to identify or analyze component parts of such music.

(c) Motivation to succeed seems to be closely related to the relevance of the task to the child.

Recommendations for Further Research

The following recommendations are made in recognition of the limitations of the present study.
(1) The Listening Test developed for this study should be further refined and administered to a larger population so that its reliability as a conceptual evaluative measure may be established.

(2) Similar listening measures using music more familiar to the child's out-of-school experiences, but containing examples of basic concepts common to all music, would perhaps more clearly determine whether or not a child actually possesses these concepts.

(3) The Listening Test developed for this study should be administered to lower-class children in small groups of four or five students to eliminate as many competitive distractions as possible.

Recommendations to Music Educators

In Chapter I, the statement was made that the inner-city child is not musically deprived, but that difficulties arise when the child's music does not conform to the musical culture of the teacher.\(^1\) If this be the situation, music educators would perhaps need to re-evaluate and re-examine their objectives and goals for the music program.

If the goal for school music experiences is to enable children to deal more effectively with the music they like and to which they respond most easily, then perhaps the answer lies in the choice of materials. Some

\(^1\)McLaurin, \textit{op. cit.}, p. 76.
forms of structure may be explored as thoroughly in a piece of popular music as in a classical orchestral composition.

If, on the other hand, the objective of a music program is to develop sensitivity to and appreciation for many kinds of music, "bridges" must be built. Through untraditional approaches, perhaps, music educators may find ways to transfer the "relevance" of familiar music to unfamiliar worthwhile music. This would require expertise on the part of the teacher, a thorough knowledge of many kinds of music, and the ability to understand a child's motivations and abilities.

Furthermore, if intellectual sensitivity to music is a goal of the music program, music educators must provide many opportunities for verbalization concerning musical events as they occur within a song or a composition. Lack of definitive vocabulary with which to communicate about musical phenomena has been cited as an almost universal problem. When a child can effectively communicate with others about music, he can more easily form value judgments based upon intelligent understanding. This would seem to be a major goal of school music experiences.
APPENDIX A

Musical Sources of Items and Correct Responses, Listening Test, Form A

Response Sheet, Listening Test, Form A
## APPENDIX A

### Music Sources of Items and Correct Responses
#### Listening Test, Form A

<table>
<thead>
<tr>
<th>Musical Source</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample A. Sullivan: Iolanthe, Overture</strong></td>
<td>louder</td>
</tr>
<tr>
<td>**Sample B. Rodgers: The King and I, &quot;The March of the Siamese Children&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>1. Pierne: Cydalise, &quot;Entrance of the Little Fauns&quot;</td>
<td>softer</td>
</tr>
<tr>
<td>2. Grieg: &quot;Wedding Day at Troldhaugen&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>3. Ravel: Mother Goose Suite, &quot;The Fairy Garden&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>4. Ravel: Mother Goose Suite, &quot;Beauty and the Beast&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>5. Handel: Water Music, &quot;Hornpipe&quot;</td>
<td>slower</td>
</tr>
<tr>
<td><strong>Sample C. Moussorgsky-Ravel: Pictures at an Exhibition, &quot;Bydlo&quot;</strong></td>
<td>lower and louder</td>
</tr>
<tr>
<td>7. Grieg: Lyric Suite, &quot;Norwegian Rustic March&quot;</td>
<td>slower and softer</td>
</tr>
<tr>
<td>8. Kabelevsky: The Comedians, &quot;Pantomine&quot;</td>
<td>higher and louder</td>
</tr>
<tr>
<td>9. Tschaikovsky: Swan Lake, &quot;Czardas&quot;</td>
<td>faster and softer</td>
</tr>
<tr>
<td>10. Copland: Appalachian Spring, &quot;Shaker Tune&quot;</td>
<td>lower and slower</td>
</tr>
<tr>
<td>11. Brahms: &quot;Hungarian Dance No. 5&quot;</td>
<td>higher and faster</td>
</tr>
<tr>
<td>12. Elgar: Wand of Youth Suite No. 1, &quot;Fairies and Giants&quot;</td>
<td>lower and louder</td>
</tr>
<tr>
<td>Musical Source</td>
<td>Correct Response</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>13. Copland: <em>Appalachian Spring</em>, &quot;Shaker Tune&quot;</td>
<td>faster</td>
</tr>
<tr>
<td>14. Verdi: <em>Aida</em>, &quot;Triumphal March&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>15. Elgar: &quot;Pomp and Circumstance No. 1&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>16. Gounod: <em>Faust Ballet Suite</em>, &quot;Waltz No. 1&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>17. Haydn: <em>Symphony No. 94</em>, Movement 2</td>
<td>softer</td>
</tr>
<tr>
<td>18. Corelli-Pinelli: <em>Suite for String Orchestra</em>, &quot;Gigue&quot;</td>
<td>slower</td>
</tr>
</tbody>
</table>
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

1. □ faster
   □ slower
   □ louder
   □ softer
   □ higher

2. □ faster
   □ slower
   □ louder
   □ softer
   □ higher

3. □ louder
   □ softer
   □ faster
   □ slower
   □ lower

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**

7. □ 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

8. □ higher and softer  
□ higher and louder  
□ higher and slower  
□ lower and louder  
□ lower and faster

11. □ lower and louder  
□ lower and faster  
□ higher and faster  
□ higher and slower  
□ lower and slower

9. □ slower and softer  
□ slower and louder  
□ faster and softer  
□ faster and louder  
□ 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.

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
APPENDIX B

Music Sources of Items and Correct Responses, Listening Test, Form B.

Response Sheet, Listening Test, Form B.
# APPENDIX B

Musical Sources and Correct Responses
Listening Test, Form B

<table>
<thead>
<tr>
<th>Musical Source</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample A. Sullivan: <em>Iolanthe</em>, &quot;Overture&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>Sample B. Rodgers: <em>The King and I</em>, &quot;The March of the Siamese Children&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>1. Pierne: <em>Cydalis</em>, &quot;Entrance of the Little Fauns&quot;</td>
<td>softer</td>
</tr>
<tr>
<td>2. Grieg: &quot;Wedding Day at Troldhaugen&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>4. Ravel: <em>Mother Goose Suite</em>, &quot;Beauty and the Beast&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>5. Handel: <em>Water Music</em>, &quot;Hornpipe&quot;</td>
<td>slower</td>
</tr>
<tr>
<td>8. Bartok: <em>Sonata for Two Pianos and Percussion</em>, Movement 1</td>
<td>faster</td>
</tr>
<tr>
<td>9. Handel: <em>Royal Fireworks Music</em>, &quot;Minuetto No. 2&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>10. Thomson: <em>Tuesday in November</em>, &quot;Fugue and Chorale on Yankee Doodle&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>11. Saint-Saens: <em>Carnival of the Animals</em>, &quot;The Swan&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>12. Schubert: <em>Quintet in C</em>, Opus 163, &quot;Scherzo&quot;</td>
<td>softer</td>
</tr>
</tbody>
</table>
13. Saint-Saëns: *Carnival of the Animals*, "Fossils"

Correct Response: lower

Sample C. Moussorgsky: *Pictures at an Exhibition*, "Bydło"


Correct Response: slower and softer

15. Kabelevsky: *The Comedians*, "Pantomine"

Correct Response: higher and louder

16. Tschaikovsky: *Swan Lake*, "Czardas"

Correct Response: faster and softer

17. Copland: *Appalachian Spring*, "Shaker Tune"

Correct Response: lower and slower

18. Brahms: "Hungarian Dance No. 5"

Correct Response: higher and faster

19. Elgar: *Wand of Youth Suite No. 1*, "Fairies and Giants"

Correct Response: lower and louder


Correct Response: higher and softer

21. Ibert: *Divertissement*, "Parade"

Correct Response: faster and louder

22. Taylor: *Through the Looking Glass*, "Garden of Live Flowers"

23. Berlioz: *Symphonie Fantastique*, Movement 1

Correct Response: slower and softer

Sample D. Schumann: "Wild Horsemen"

25. Copland: *Appalachian Spring*, "Shaker Tune"

Correct Response: faster

26. Verdi: *Aida*, "Triumphant March"

Correct Response: higher

27. Elgar: "Pomp and Circumstance No. 1"

Correct Response: louder

28. Gounod: *Faust Ballet Suite*, "Waltz No. 1"

Correct Response: lower

29. Haydn: *Symphony No. 94*, Movement 2

Correct Response: softer

30. Corelli-Pinelli: *Suite for String Orchestra*, "Gigue"

Correct Response: slower
<table>
<thead>
<tr>
<th>Musical Source</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>31. Dixon-Henderson: &quot;Bye, Bye, Blackbird&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>32. Thomson: <em>The Plow that Broke the Plains</em>, &quot;Cattle&quot;</td>
<td>slower</td>
</tr>
</tbody>
</table>
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, put a check mark beside the answer you think is right. Check only one answer for each example.

<table>
<thead>
<tr>
<th>SAMPLE A</th>
<th></th>
<th>SAMPLE B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>_______faster</td>
<td></td>
<td>louder</td>
<td></td>
</tr>
<tr>
<td>_______slower</td>
<td></td>
<td>_______higher</td>
<td></td>
</tr>
<tr>
<td>_______louder</td>
<td></td>
<td>_______lower</td>
<td></td>
</tr>
<tr>
<td>_______softer</td>
<td></td>
<td>_______faster</td>
<td></td>
</tr>
<tr>
<td>_______higher</td>
<td></td>
<td>_______lower</td>
<td></td>
</tr>
</tbody>
</table>

|   |   |   |   |
| 1. | _______faster |   | _______faster |
|   | _______slower |   | _______slower |
|   | _______louder |   | _______louder |
|   | _______softer |   | _______softer |
|   | _______higher |   | _______higher |

|   |   |   |   |
| 2. | _______faster |   | _______faster |
|   | _______slower |   | _______slower |
|   | _______louder |   | _______louder |
|   | _______higher |   | _______higher |

|   |   |   |   |
| 3. | _______louder |   | _______louder |
|   | _______softer |   | _______softer |
|   | _______faster |   | _______faster |
|   | _______slower |   | _______slower |

|   |   |   |   |
| 4. | _______louder |   | _______lower |
|   | _______softer |   | _______lower |
|   | _______higher |   | _______lower |

|   |   |   |   |
| 5. | _______softer |   | _______louder |
|   | _______higher |   | _______louder |
|   | _______lower |   | _______lower |

|   |   |   |   |
| 6. | _______faster |   | _______faster |
|   | _______slower |   | _______slower |
|   | _______louder |   | _______louder |
|   | _______softer |   | _______softer |
|   | _______higher |   | _______higher |

|   |   |   |   |
| 7. | _______faster |   | _______lower |
|   | _______slower |   | _______lower |
|   | _______louder |   | _______louder |
|   | _______softer |   | _______softer |

|   |   |   |   |
| 8. | _______softer |   | _______lower |
|   | _______slower |   | _______lower |
|   | _______faster |   | _______faster |
|   | _______higher |   | _______higher |

|   |   |   |   |
| 9. | _______faster |   | _______lower |
|   | _______slower |   | _______lower |
|   | _______louder |   | _______louder |
|   | _______softer |   | _______softer |

|   |   |   |   |
| 10. | _______faster |   | _______louder |
|   | _______slower |   | _______slower |
|   | _______higher |   | _______higher |

|   |   |   |   |
| 11. | _______louder |   | _______softer |
|   | _______lower |   | _______lower |
|   | _______softer |   | _______softer |
|   | _______higher |   | _______higher |

|   |   |   |   |
| 12. | _______faster |   | _______lower |
|   | _______slower |   | _______slower |
|   | _______louder |   | _______louder |
|   | _______softer |   | _______softer |

|   |   |   |   |
| 13. | _______louder |   | _______lower |
|   | _______softer |   | _______lower |
|   | _______higher |   | _______higher |
|   | _______faster |   | _______faster |
In this section each answer has two parts. For example, the music may get both higher and slower, or perhaps faster and louder. Choose the answer that has both parts right, but check only one answer for each example.

<table>
<thead>
<tr>
<th>SAMPLE C</th>
<th>19. lower and softer</th>
</tr>
</thead>
<tbody>
<tr>
<td>______ slower and softer</td>
<td>______ lower and louder</td>
</tr>
<tr>
<td>______ slower and louder</td>
<td>______ higher and softer</td>
</tr>
<tr>
<td>______ faster and softer</td>
<td>______ slower and softer</td>
</tr>
<tr>
<td>______ higher and louder</td>
<td>______ higher and louder</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. slower and softer</th>
<th>20. higher and louder</th>
</tr>
</thead>
<tbody>
<tr>
<td>______ slower and lower</td>
<td>______ higher and slower</td>
</tr>
<tr>
<td>______ faster and softer</td>
<td>______ lower and softer</td>
</tr>
<tr>
<td>______ faster and louder</td>
<td>______ lower and faster</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>15. higher and softer</th>
<th>21. slower and softer</th>
</tr>
</thead>
<tbody>
<tr>
<td>______ higher and louder</td>
<td>______ slower and louder</td>
</tr>
<tr>
<td>______ higher and slower</td>
<td>______ faster and slower</td>
</tr>
<tr>
<td>______ lower and louder</td>
<td>______ faster and louder</td>
</tr>
<tr>
<td>______ lower and faster</td>
<td>______ higher and louder</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>16. slower and softer</th>
<th>22. slower and softer</th>
</tr>
</thead>
<tbody>
<tr>
<td>______ slower and louder</td>
<td>______ slower and lower</td>
</tr>
<tr>
<td>______ faster and softer</td>
<td>______ faster and slower</td>
</tr>
<tr>
<td>______ faster and louder</td>
<td>______ higher and softer</td>
</tr>
<tr>
<td>______ higher and softer</td>
<td>______ higher and softer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>17. higher and softer</th>
<th>23. higher and softer</th>
</tr>
</thead>
<tbody>
<tr>
<td>______ higher and louder</td>
<td>______ lower and louder</td>
</tr>
<tr>
<td>______ higher and slower</td>
<td>______ higher and slower</td>
</tr>
<tr>
<td>______ lower and slower</td>
<td>______ lower and faster</td>
</tr>
<tr>
<td>______ lower and faster</td>
<td>______ lower and faster</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>18. lower and louder</th>
<th>24. slower and louder</th>
</tr>
</thead>
<tbody>
<tr>
<td>______ lower and faster</td>
<td>______ slower and slower</td>
</tr>
<tr>
<td>______ higher and faster</td>
<td>______ faster and slower</td>
</tr>
<tr>
<td>______ higher and slower</td>
<td>______ faster and louder</td>
</tr>
<tr>
<td>______ louder and slower</td>
<td>______ higher and louder</td>
</tr>
</tbody>
</table>
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 part. Check the answer which best describes this difference. Check only one answer for each example. Listen carefully as these are played, because they will not be repeated.

SAMPLE D

- faster
- slower
- higher
- lower
- softer

25. louder
- higher
- softer
- faster
- slower

26. lower
- higher
- slower
- faster
- softer

27. louder
- slower
- faster
- softer
- lower

28. higher
- lower
- faster
- slower
- softer
APPENDIX C

Musical Sources of Items and Correct Responses, Listening Test, Form C

Response Sheet, Listening Test, Form C
### APPENDIX C

Musical Sources of Items and Correct Responses
Listening Test, Form C

<table>
<thead>
<tr>
<th>Musical Source</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample A. Sullivan: Iolanthe, &quot;Overture&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>Sample B. Rodgers: The King and I, &quot;The March of the Siamese Children&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>1. Pierne: Cydalise, &quot;Entrance of the Little Fauns&quot;</td>
<td>softer</td>
</tr>
<tr>
<td>2. Grieg: &quot;Wedding Day at Troldhaugen&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>3. Ravel: Mother Goose Suite, &quot;Beauty and the Beast&quot;</td>
<td>lower and louder</td>
</tr>
<tr>
<td>4. Handel: Water Music, &quot;Hornpipe&quot;</td>
<td>slower</td>
</tr>
<tr>
<td>5. Shostakovich: Ballet Suite No. 1, &quot;Pizzicato Polka&quot;</td>
<td>faster</td>
</tr>
<tr>
<td>6. Thomson: The Plow that Broke the Plains, &quot;Cattle&quot;</td>
<td>softer</td>
</tr>
<tr>
<td>7. Bartok: Sonata for Two Pianos and Percussion, Movement 1</td>
<td>faster</td>
</tr>
<tr>
<td>8. Handel: Royal Fireworks Music, &quot;Minuetto No. 2&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>9. Thomson: Tuesday in November, &quot;Fugue and Chorale on Yankee Doodle&quot;</td>
<td>higher</td>
</tr>
<tr>
<td>10. Schubert: Quintet in C, Opus 163, &quot;Scherzo&quot;</td>
<td>softer</td>
</tr>
<tr>
<td>11. Saint-Saens: Carnival of the Animals, &quot;Fossils&quot;</td>
<td>lower</td>
</tr>
</tbody>
</table>

Sample C. Moussorgsky: Pictures at an Exhibition, "Bydlo"
<table>
<thead>
<tr>
<th>Musical Source</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Grieg: Lyric Suite, &quot;Norwegian Rustic March&quot;</td>
<td>slower and softer</td>
</tr>
<tr>
<td>13. Tschaikovsky: Swan Lake, &quot;Czardas&quot;</td>
<td>faster and softer</td>
</tr>
<tr>
<td>14. Elgar: Wand of Youth Suite No. 1, &quot;Fairies and Giants&quot;</td>
<td>lower and louder</td>
</tr>
<tr>
<td>15. Bizet: Carmen Suite No. 2, &quot;Changing of the Guard&quot;</td>
<td>higher and softer</td>
</tr>
<tr>
<td>16. Ibert: Divertissement, &quot;Parade&quot;</td>
<td>faster and softer</td>
</tr>
<tr>
<td>17. Rossini: William Tell Overture</td>
<td>faster and softer</td>
</tr>
<tr>
<td>18. Taylor: Through the Looking Glass, &quot;Garden of Live Flowers&quot;</td>
<td>lower and slower</td>
</tr>
<tr>
<td>19. Berlioz: Symphonie Fantastique, Movement 1</td>
<td>slower and softer</td>
</tr>
<tr>
<td>Sample D. Schumann: &quot;Wild Horsemen&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>20. Copland: Appalachian Spring, &quot;Shaker Tune&quot;</td>
<td>faster</td>
</tr>
<tr>
<td>21. Elgar: &quot;Pomp and Circumstance No. 1&quot;</td>
<td>louder</td>
</tr>
<tr>
<td>22. Gounod: Faust Ballet Suite, &quot;Waltz No. 1&quot;</td>
<td>lower</td>
</tr>
<tr>
<td>23. Haydn: Symphony No. 94, Movement 2</td>
<td>softer</td>
</tr>
<tr>
<td>24. Thomson: The Plow that Broke the Plains, &quot;Cattle&quot;</td>
<td>slower</td>
</tr>
</tbody>
</table>
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, put a check mark beside the answer you think is right. Check only one answer for each example.

SAMPLE A

<table>
<thead>
<tr>
<th>1. faster</th>
<th>2. faster</th>
<th>3. louder</th>
</tr>
</thead>
<tbody>
<tr>
<td>slower</td>
<td>slower</td>
<td>softer</td>
</tr>
<tr>
<td>louder</td>
<td>louder</td>
<td>higher</td>
</tr>
<tr>
<td>softer</td>
<td>higher</td>
<td>slower</td>
</tr>
</tbody>
</table>

SAMPLE B

<table>
<thead>
<tr>
<th>4. softer</th>
<th>5. faster</th>
<th>6. faster</th>
</tr>
</thead>
<tbody>
<tr>
<td>higher</td>
<td>slower</td>
<td>slower</td>
</tr>
<tr>
<td>lower</td>
<td>louder</td>
<td>louder</td>
</tr>
<tr>
<td>faster</td>
<td>softer</td>
<td>higher</td>
</tr>
<tr>
<td>slower</td>
<td>lower</td>
<td>lower</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. softer</th>
</tr>
</thead>
<tbody>
<tr>
<td>slower</td>
</tr>
<tr>
<td>faster</td>
</tr>
<tr>
<td>higher</td>
</tr>
<tr>
<td>lower</td>
</tr>
</tbody>
</table>
In this section each answer has two parts. For example, the music may get both higher and slower, or perhaps faster and louder. Choose the answer that has both parts right, but check only one answer for each example.

<table>
<thead>
<tr>
<th>SAMPLE C</th>
<th>16.</th>
<th>17.</th>
<th>18.</th>
<th>19.</th>
</tr>
</thead>
<tbody>
<tr>
<td>slower and softer</td>
<td>slower and softer</td>
<td>slower and softer</td>
<td>slower and louder</td>
<td>slower and louder</td>
</tr>
<tr>
<td>slower and louder</td>
<td>faster and softer</td>
<td>faster and louder</td>
<td>faster and louder</td>
<td>faster and softer</td>
</tr>
<tr>
<td>faster and louder</td>
<td>higher and louder</td>
<td>higher and louder</td>
<td>higher and louder</td>
<td>higher and louder</td>
</tr>
<tr>
<td>slower and softer</td>
<td>slower and softer</td>
<td>slower and louder</td>
<td>lower and louder</td>
<td>slower and louder</td>
</tr>
<tr>
<td>slower and louder</td>
<td>faster and softer</td>
<td>faster and louder</td>
<td>higher and slower</td>
<td>lower and slower</td>
</tr>
<tr>
<td>faster and softer</td>
<td>higher and slower</td>
<td>higher and softer</td>
<td>lower and faster</td>
<td>lower and faster</td>
</tr>
</tbody>
</table>
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 part. Check the answer which best describes this difference. Check only one answer for each example. Listen carefully as these are played, because they will not be repeated.

SAMPLE D
- faster
- slower
- higher
- lower
- softer

20.
- louder
- higher
- softer
- faster
- slower

21.
- louder
- slower
- faster
- softer
- lower

22.
- faster
- slower
- higher
- lower
- softer

23.
- lower
- slower
- faster
- louder
- softer

24.
- lower
- faster
- slower
- louder
- higher
APPENDIX D

OSU Item Analysis Program
O.S.U. ITEM ANALYSIS PROGRAM

I. Purpose
To provide a rapid and efficient means of scoring and item-analyzing objective, single-part tests in accordance with standard analysis procedures.

II. Analyses Performed

1. Analysis performed on master control card.

2. A score analysis which gives a score list, frequency distribution, cumulative frequency, and percentiles.

III. Item statistics

a. item difficulty
b. item discrimination (phi coefficient and point biserial r)
c. number in and proportion of total group answering item correctly
d. alternative count for upper and lower groups of subjects taking the test (up to 9 alternatives).

IV. Summary statistics

a. total number of persons taking the test
b. number of persons in each of upper and lower groups
c. number of items on the test
d. mean scores for upper, lower, and total groups
e. median score for total group
f. standard deviation for total group

g. reliability coefficient and standard error by one or all of the following formulae:
   1. Kuder-Richardson #20
   2. Kuder-Richardson #21
   3. Odd-Even split corrected by the Spearman-Brown Prophecy formula

V. A card containing the ID, test score, standardized score, and percentile for each individual (optional).
6. A listing containing test ID, student ID and name, total score, standardized score and percentile for each individual (optional).

Note: upper and lower groups may be chosen by means of an external criterion (optional).

C. Program Language and Requirements

This is a SCA\textsuperscript{4}RAN program in four segments and will time-share. The program at present will accommodate up to 2400 subjects and 120 items. These numbers may be extended by changing the parameters card, but the number of items must be a multiple of 60. For example, if the program must handle 150 items, the parameter must be changed from 120 to 180. Any extension of parameters or use of the COPIES job specification will eliminate the time-sharing feature.

D. Deck Setup

1) Master Control Card (required)

<table>
<thead>
<tr>
<th>Col.</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>blank</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>indicator - is external criterion deck provided?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 no</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 yes</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>blank</td>
<td></td>
</tr>
<tr>
<td>5-7</td>
<td>number of items on external test (if provided)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>blank</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>no. alternatives per item on ext. test (if provided)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>blank</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>number of key cards for ext. test (if provided)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>blank</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>number of cards per subject on ext. test (if provided)</td>
<td></td>
</tr>
<tr>
<td>14-26</td>
<td>percentage of total to constitute upper group (of form 000.xxx)</td>
<td></td>
</tr>
</tbody>
</table>
percentage of total to constitute lower group (of form 000.xxx)
*for example, 27.5% would be punched 000.275

28 blank

29 indicator - is name deck provided?
0 no
1 yes

30 blank

31 indicator - on external test deck, are total scores given (if provided)
0 no, totals not given - deck must be scored
1 yes

32 blank

33 indicator - on external test deck, how is subject data punched?
0 keypunched
1 IBM 1230 optical scorer special card code
2 IBM 1230 optical scorer standard card code

34 blank

35 indicator - on test to be analyzed, how is regular subject data punched?
0 keypunched
1 IBM 1230 special card code
2 IBM 1230 standard card code

II. External Criterion Test Deck

(optional depending on col. 3 of master control card)

a. key card(s) - must be omitted if col. 31 above punched with a 1
- same format as described later for regular data
- key cards if col. 31 above punched with a 0

b. data cards - if 1 in col. 31 above, first 11 columns same as for regular data described later; col. 12-14 have score
- if 0 in col. 31 above, same format as described later
c. terminal card

Col.  i  6
   "0  blank

2. Name Dec (optional depending on col. 29 of master control card above)

a. one card for each student number appearing in the data decks
(these must be sorted in increasing numerical order by student ID number)

Col.  1  blank
   2-3  ignored by program
   4-10  student ID number
   11-40  subject's name

b. terminal card

Col.  i  6
   2-30  blank

4. Regular Data (required)

a. Title card(s)

This title (up to 3 lines of 66 characters or less per line) appears
at appropriate places in the output listing.

Col.  0
   2-4  blank
   5  title card sequence number (≤ 3)
   6  blank
   7-72  desired line of title
   73-80  not used

b. parameter card

Certain parameters defining the test and desired options are supplied
the program by this card.
Col. 1
blank

(2) number of items on test

(3) blank

(4) number of persons taking test (not used by program)

(5) blank

(6) number of alternatives per item (up to 9)

(7) blank

(8) item discrimination coefficient (not used by program -- both phi coefficient and point biserial r are arbitrarily computed)

(9) blank

(10) reliability coefficient desired

0 KR21

1 KR20

2 Odd-Even

3 all of above

blank

(12) raw score listing options

blank nothing

0 punch cards only

1 punch cards and list

2 list only

(13) blank

(14) number of key cards

(15) blank

(16) number of data cards per subject (1 or 2)

(17) blank

(18) new standard deviation (of form xxx.xxx) see p. 7, #8

(i.e., $010.000$)

(19) blank

(20) new mean (of form xxx.xxx) see p. 7, #8

(i.e., $050.000$)

(21) ignored by program
c. key card(s) for scoring the test

Col.  1  2

2  key card sequence number (1 or 2)
3-4  test ID number
5  form number
6-11  ignored by program
12-71  key or correct answers for each item
72-80  not used

Key cards must not be punched in either of the IBM 1230 card codes.

d. data cards

Containing the responses to each item on the test by each individual taking the test.

Col.  1  3

2  card sequence number (1 or 2)
3  ignored by program
* 4-10  student ID number
11  ignored by program
** 12-71  subject's responses for each item
72-80  not used

**Col. 4-10 must be a unique 7 digit number for each student. Blanks are permitted in high-order (left) columns if fewer than 7 digits are given. These cards should be sorted in increasing numerical order by student ID number. If data is out of order, it will be sorted by the program; this is costly in execution time, however.

**Responses are normally punched 1 per column, making 60 responses per card. If data is punched by the IBM 1230-534 scorer in its special card code, there will be 120 responses per card.

e. terminal card

Col.  1  5

2-30  blank
5. **End of Run Card**

Col. 1 4

3.80 blank

***Any number of sets (1. - 4.) may precede the End Run card. The End Run card appears only once at the very end -- not between individual forms.

E. **Method**

1. **Item Difficulty**

The difficulty is found by dividing the total number getting the item correct by the total number taking the test.

2. **Phi Coefficient**

\[
\phi = \frac{a \delta - \beta \gamma}{\sqrt{p \cdot q \cdot p' \cdot q'}}
\]

where

- \(a\) is the ratio of the upper group answering the item correctly to the total number in the two groups.
- \(\beta\) is the ratio of the lower group answering the item correctly to the total number in the two groups.
- \(\gamma\) is the ratio of the upper group answering the item incorrectly to the total number in the two groups.
- \(\delta\) is the ratio of the lower group answering the item incorrectly to the total number in the two groups.

\[
p = a + \beta
\]

\[
q = \gamma + \delta
\]

\[
p' = a + \gamma
\]

\[
q' = \beta + \delta
\]

3. **Point Biserial \(r\) Coefficient**

\[
\rho_{\text{bis}} = \frac{M_p - M_t}{s} \cdot \sqrt{p/q}
\]
F. Method

1. Item Difficulty

The difficulty is found by dividing the total number getting the item correct by the total number taking the test.

2. Phi Coefficient

\[
\phi = \frac{a \delta - \beta \gamma}{\sqrt{p \cdot q \cdot p' \cdot q'}}
\]

where
- \(a\) is the ratio of the upper group answering the item correctly to the total number in the two groups.
- \(\beta\) is the ratio of the lower group answering the item correctly to the total number in the two groups.
- \(\gamma\) is the ratio of the upper group answering the item incorrectly to the total number in the two groups.
- \(\delta\) is the ratio of the lower group answering the item incorrectly to the total number in the two groups.

\[
p = a + \beta
\]
\[
q = \gamma + \delta
\]
\[
p' = a + \gamma
\]
\[
q' = \beta + \delta
\]

3. Point Biserial \(r\) Coefficient

\[
r_{pbis} = \frac{M_p - M_q}{s} \cdot \sqrt{p/q}
\]
where $M_p$ is the mean test score of those getting the item correct

$M_t$ is the mean test score of the total group

$p$ is the number of students getting the item correct

$q$ is the number of students getting the item wrong

$s$ is the standard deviation

4. **D Coefficient**

$$D = p - q$$

where $p$ is the ratio of people in the upper group answering the item correctly to the total number of people in the upper group

$q$ is the ratio of people in the lower group answering the item correctly to the total number of people in the lower group

5. **Upper and Lower Groups**

The percentages of the total which are to constitute the upper and lower groups are specified by the user on the master control card. Groups are selected on the basis of high (or low) test scores either on an external criterion test or on the test being analyzed. This is specified in column 3 of the master control card.

6. **Mean Score**

The mean score for a given group (upper, lower, or total) is the sum of scores of the group divided by the total number in the group.
7. Median Score

\[ MD = L + \frac{i(N/2 - F)}{f} \]

where
- \( MD \) is the Median
- \( L \) is the lower limit of the Median interval
- \( i \) is the length of the Median interval
- \( f \) is the frequency for the Median interval
- \( F \) is the cumulative frequency for all intervals below the Median interval
- \( N \) is the total number of subjects taking the test

8. Standard Deviation

\[ \sigma = \sqrt{\frac{N\Sigma x^2 - (\Sigma x)^2}{N(N - 1)}} \]

where
- \( \sigma \) is the standard deviation of the test
- \( x \) is the individual score
- \( N \) is the total number of subjects taking the test

9. Standard Score (computed for each individual subject)

\[ \left( \frac{x - \bar{x}}{\sigma} \right) \cdot \mu + \mu \]

where
- \( x \) is the individual score
- \( \bar{x} \) is the mean score
- \( \sigma \) is the standard deviation
- \( \sigma \) is the new standard deviation specified on regular data parameter card
- \( \mu \) is the new mean specified on the regular data parameter card
10. Reliability Coefficient

Reliability coefficients may be computed from one or all of the following formulae:

a. Kuder-Richardson formula #20

\[
\hat{r}_{tt} = \frac{k}{k-1} \left( \frac{S^2 - \sum p \cdot q}{S^2} \right)
\]

where \( \hat{r}_{tt} \) is the estimate of reliability
\( k \) is the number of items in the test
\( s \) is the standard deviation of the test
\( p \) is the proportion passing a particular item
\( q \) is the proportion failing the same item

b. Kuder-Richardson formula #21

\[
\hat{r}_{tt} = \frac{k}{k-1} \left[ 1 - \bar{x} \left( 1 - \frac{\bar{x}}{k} \right) \right]
\]

where \( \bar{x} \) is the mean score of the group
other symbols have same meaning as in a. above

c. Odd-Even split

\[
\hat{r}_{tt} = \frac{N \sum \bar{O} \cdot E - (\sum \bar{O}) \cdot (\sum E)}{\sqrt{[N \sum \bar{O}^2 - (\sum \bar{O})^2] \cdot [N \sum E^2 - (\sum E)^2]}}
\]
where $\bar{U}$ is the off-number score
$E$ is the even number score
$N$ is the number of pairs of scores

This correlation yields a relationship between two tests of $1/2$ the length of the true test and is corrected for attenuation by the Spearman-Brown Prophecy formula:

$$r = \frac{2r_{tt}}{1 + r_{tt}}$$

... Standard Error

For each of the above reliability coefficients above, a standard error is computed according to the following formula:

$$Err = s \sqrt{1-r}$$

where $s$ is the standard deviation
$r$ is the appropriate reliability coefficient
APPENDIX E

BMD02R Stepwise Regression Analysis Program
1. GENERAL DESCRIPTION

a. This program computes a sequence of multiple linear regression equations in a stepwise manner. At each step one variable is added to the regression equation. The variable added is the one which makes the greatest reduction in the error sum of squares. Equivalently it is the variable which has highest partial correlation with the dependent variable partialled on the variables which have already been added; and equivalently it is the variable which, if it were added, would have the highest F value. In addition, variables can be forced into the regression equation. Non-forced variables are automatically removed when their F values become too low. Regression equations with or without the regression intercept may be selected.

b. Output from this program includes:

(1) At each step:
   (a) Multiple R
   (b) Standard error of estimate
   (c) Analysis-of-variance table
   (d) For variables in the equation:
       1. Regression coefficient
       2. Standard error
       3. F to remove
   (e) For variables not in the equation:
       1. Tolerance
       2. Partial correlation coefficient
       3. F to enter

(2) Optional output prior to performing regression:
   (f) Means and standard deviations
   (g) Covariance matrix
   (h) Correlation matrix

(3) Optional output after performing regression:
   (i) List of residuals
   (j) Plots of residuals vs. input variables
   (k) Summary table
c. Limitations per problem:

(1) \( p \), number of original variables \( (2 \leq p \leq 80) \)
(2) \( q \), number of variables added by transgeneration \( (-9 \leq q \leq 78) \)
(3) \( p+q \), total number of variables \( (2 \leq p+q \leq 80) \)
(4) \( s \), number of Sub-problem Cards \( (1 \leq s \leq 99) \)
(5) \( k \), number of Variable Format Cards \( (1 \leq k \leq 10) \)
(6) \( i \), number of variables to be plotted \( (0 \leq i \leq 30) \)
(7) \( n \), number of cases \( (1 \leq n \leq 9999) \)
(8) \( m \), number of Transgeneration Cards \( (0 \leq m \leq 99) \)

d. Estimation of running time and output pages per problem:

Number of seconds \( = \frac{(p+q) \times n}{100} \) (for IBM 7094)

Number of pages \( = \frac{\text{no. of steps } [23 + \frac{3}{4}(p+q)]}{56} + 5 \text{ per sub-problem} \)

e. This program allows transgeneration of the variables. Codes 01-17 and 20-24 of the transgeneration list may be used.

2. ORDER OF CARDS IN JOB DECK

Cards indicated by letters enclosed in parentheses are optional. All other cards must be included in the order shown.

- a. System Cards [Introduction, IV]

- b. Problem Card

- (c.) Transgeneration Card(s) [Introduction, III-B]

- (d.) Labels Card(s) [Introduction, III-A]

- (e.) F-type Variable Format Cards [Introduction, III-C]

- (f.) DATA INPUT Cards
(Place data input deck here if data input is from cards.) [Introduction, II-C]

- g. Sub-problem Card(s)

- (h.) Control-Delete Card(s)

- (i.) Index-Plot Card(s)
g. through (i.) may be repeated as many as 99 times in each problem; b. through (i.) may be repeated as often as desired.

j. Finish Card 

3. CARD PREPARATION (SPECIFIC FOR THIS PROGRAM)

Preparation of the cards listed below is specific for this program. All other cards listed in the preceding section are prepared according to instructions in the Introduction.

b. Problem Card (One Problem Card for each problem)

Col. 1-6 PRØBLM (Mandatory)

Col. 10-15 Alphanumeric problem name

Col. 17-20 Sample size \( (1 \leq n \leq 9999) \)

Col. 24, 25 Number of original variables \( (2 \leq p \leq 80) \)

Col. 29, 30 Number of Transgeneration Cards \( (0 \leq m \leq 99) \)

Col. 34, 35 Number of variables added by transgeneration \( (-9 \leq q \leq 78) \)

Col. 39, 40 Tape number if data is on tape \( (=/= \text{logical } 2) \); otherwise, leave blank

Col. 44, 45 Number of Sub-problem Cards \( (1 \leq s \leq 99) \)

Col. 48, 49 Number of variables labeled on Labels Cards. Leave blank if Labels Cards are not used.

Col. 51-53 YES If means and standard deviations are to be printed; otherwise, leave blank.

Col. 55-57 YES If covariance matrix is to be printed; otherwise, leave blank.

Col. 59-61 YES If correlation matrix is to be printed; otherwise, leave blank.
Example of Job Deck Set-up:

- **j.** FINISH
  - Finish Card
- **g.** through (i.) repeated as many as 99 times if desired for each problem
- **i.** IDXPLT
  - Index-Plot Card(s)
- **h.** CÔNDEL
  - Control-Deletion Card(s)
- **g.** SUBPRØ
  - Sub-problem Card
- **f.** Data Input Deck
- **e.** F-type Variable Format Card(s)
- **d.** LABELS
  - Labels Card(s)
- **c.** TRNGEN
  - Standard Transgeneration Card(s)
- **b.** PROBLM
  - Problem Card
- **a.** $ID
  - System Card(s)
Col. 63-65  YES  If zero regression intercept is desired; otherwise, leave blank.

Col. 68, 69  NO  If tape specified in Columns 39, 40 is not to be rewound before this problem; leave blank if Columns 39, 40 are blank, or if tape rewind is desired.

Col. 71, 72  Number of F-type Variable Format Cards ($1 \leq k \leq 10$)

**g. Sub-problem Card**

Col. 1-6  SUBPRØ  (Mandatory)

Col. 9, 10  Number of the dependent variable.

Col. 13-15  Maximum number of steps. This will be $2(p+q)$ if left blank.

Col. 20-25  F-level for inclusion. This will be 0.01 if left blank.

Col. 30-35  F-level for deletion. This will be 0.005 if left blank.

Col. 40-45  Tolerance level. This will be 0.001 if left blank.

Col. 49, 50  Number of variables on the Index-Plot Card ($0 \leq i \leq 30$)

Col. 53-55  YES  If Control-Delete Cards are included.

Col. 58-60  YES  If list of residuals is to be printed.

Col. 63-65  YES  If summary table is to be printed.

**h. Control-Delete Card**

Col. 1-6  CØNDEL  (Mandatory)

Col. 7  Control value* for first variable

Col. 8  Control value* for second variable

...  

Col. 72  Control value* for 66th variable
If there are more than 66 variables, continue on another card of the same form, until \( p+q \) variables have been specified.

The variable numbers above refer to variables after trans-generation.

*CONTROL VALUES*

1. Delete variable (or dependent variable)
2. Free variable
3. Low-level forced variable
   ...
9. High-level forced variable

If no Control-Delete Cards are included, or if a field is left blank on the Control-Delete Cards included in the deck, the value 2 will be assigned if the variable is not the dependent variable and the value 1 assigned if it is the dependent variable.

(i.) Index-Plot Card

Variables specified on this card are plotted against the residuals.

<table>
<thead>
<tr>
<th>Col. 1-6</th>
<th>IDXPLT</th>
<th>(Mandatory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col. 7, 8</td>
<td>First variable to be plotted</td>
<td></td>
</tr>
<tr>
<td>Col. 9, 10</td>
<td>Second variable to be plotted</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col. 65, 66</td>
<td>30th variable to be plotted</td>
<td></td>
</tr>
</tbody>
</table>

No more than 30 variables may be plotted per sub-problem.

Variables specified refer to the original data after trans-generation.

4. **COMPUTATIONAL PROCEDURE**

Step 1. The data are read and transgenerated (see Introduction, Section III-B). Let \( p \) denote the number of variables after transgeneration, \( n \) the number of cases and \( x_{ij} \) the value of the \( j \)th variable, after transgeneration, for the \( i \)th case. The means
are computed and, if called for, printed. If a zero regression intercept is not requested on the Problem Card, the matrix $A$

\[
A_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad i, j = 1, \ldots, p
\]

is computed. If a zero regression intercept is requested, the matrix $A$

\[
A_{ij} = \sum_{k=1}^{n} x_{ki} x_{kj} \quad i, j = 1, \ldots, p
\]

is computed instead.

**Step 2.** The covariances, standard deviations, and correlations

\[
s_{ij} = \frac{1}{m} a_{ij}; \quad i, j = 1, \ldots, p
\]

\[
m = \begin{cases} 
  n, & \text{if zero regression intercept is requested} \\
  n-1, & \text{otherwise}
\end{cases}
\]

\[
s_i = \sqrt{s_{ii}} \quad i = 1, \ldots, p
\]

\[
r_{ij} = \frac{s_{ij}}{s_is_j} \quad i, j = 1, \ldots, p
\]

are computed and, if called for, printed. It should be noted that if the zero regression intercept option is chosen, these statistics will not be centered about the mean. A similar statement applies to all the computations which follow.

**Step 3.** At each step in the stepwise regression procedure the variables $x_1, \ldots, x_p$ are divided into two disjoint sets:

- $x_1, \ldots, x_q$ : The independent variables in the regression equation.
- $x_{i1}, \ldots, x_{iq}$ : The remaining variables including the dependent variable $y = x_{i1}$.

For purposes of exposition, assume that $x_{i1}, \ldots, x_{iq}$ are the first $q$ variables $x_1, \ldots, x_q$. The regression
equation at a typical step then has the form
\[ y = \alpha + \beta_1 x_1 + \ldots + \beta_p x_p + e \]

Let
\[ A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \]

where \( A_{11} \) is \( q \times q \)

be a partition of the matrix \( A \) from Step 1; let
\[ B = \begin{bmatrix} A_{11}^{-1} & A_{11}^{-1} A_{12} \\ A_{21} A_{11}^{-1} & A_{22} - A_{21} A_{11}^{-1} A_{12} \end{bmatrix} \]

let \( m = n-1 \) if a zero regression intercept is not requested; and let \( m = n \) if it is requested. For each step in the stepwise procedure the following are computed and printed:

The residual degrees of freedom, sum of squares, and mean square.
\[ df = m - q, \quad SS = b_{dd}, \quad MS = SS/df \]

The regression degrees of freedom, sum of squares, mean square, and \( F \) value.
\[ rdf = q, \quad RSS = a_{dd} - b_{dd} \]
\[ RMS = RSS/rdf, \quad F = RMS/MS \]

The standard error of estimate and multiple correlation coefficient.
\[ S = \sqrt{MS}, \quad R = \sqrt{RSS/a_{dd}} \]

For each independent variable \( x_i \) in the regression equation, the following are computed and printed:

The regression coefficient, its standard error, and \( F \) value.
\[ \beta_i = b_{i1}, \quad S_i = b_{ii}^{1/2} S, \quad F_i = (\beta_i/S_i)^2 \]
If a zero regression intercept is not requested, the intercept $\alpha$ is computed.

$$\alpha = \sum_{i=1}^{q} \beta_i x_i$$

For each independent variable $x_i$ not in the regression equation, the following are computed and printed:

The tolerance level, partial correlation, coefficient, and $F$ value.

$$T_i = b_{id}/a_{ii}, \quad R_i = \frac{b_{id}}{\sqrt{b_{ii}b_{dd}}}, \quad F_i = \frac{b_{id}^2 (m-q-1)}{b_{ii}b_{dd} - b_{id}^2}$$

**Step 4.** To move from one step to the next, an independent variable is added to or removed from the regression equation according to the following three rules:

1. If there are one or more independent variables in the regression equation whose control value, as specified by the Control-Delete Card, is 2 (i.e., a free variable) and whose $F$ value is less than the "$F$-to-remove" value specified on the Subproblem Card, the one with the smallest $F$ value will be removed.

2. If no variable is removed by (1) and there are one or more independent variables not in the regression equation which pass the tolerance test and have control values of 3 or more (i.e., forced variables), the one which has the highest control value and the highest $F$ value among all with the same control value will be added.

An independent variable $x_i$ not in the regression equation is said to pass the tolerance test if its tolerance value $T_i$ is greater than or equal to the "minimum tolerance value" specified on the Subproblem Card.
(3) If no variable is removed by (1) or added by (2) and there are one or more independent variables not in the regression equation which pass the tolerance test, have a control value of 2 (i.e., a free variable), and an F value greater than or equal to the "F-to-enter" value specified on the Sub-problem Card, the one with the highest F value will be added.

If no variable is added or removed by (1), (2), or (3), the stepwise procedure is terminated.

Step 5. If a list of residuals is called for on the Sub-problem Card, the residuals

\[ r_i = y_i - \alpha - \sum_j \beta_j x_{ij} \quad i = 1, \ldots, n \]

are computed and printed. The summation is over all indices \( j \) of independent variables \( x_j \) in the regression equation at the last step.

For each variable \( x_j \) specified on the Index-Plot Card, the points

\( (r_i, x_{ij}) \quad i = 1, \ldots, n \)

are computed and plotted.

5. MACHINE PROCEDURE

The Computational Procedure section explains in terms of the input data what is computed; however, it offers few details on how the machine accomplishes the computation. The purpose of this section is to present more detailed information about the machine procedure. It is written in the imperative mood to conform with the commands presented by the program to the machine.

Main Program

1. Read a Problem Card. Read the data and transgenerate. Let

\[ x_{ij}; i = 1, \ldots n; j = 1, \ldots p \]

denote the \( i^{th} \) case value of the \( j^{th} \) variable after transgeneration. If a regression intercept is called for, compute a matrix \( A \) and
vector B as follows. Set
\[ a_{ij} = 0 \quad 1 \leq i \leq j \leq p \]
\[ b_i = x_{1i} \quad i = 1, \ldots, p \]
and apply the recursion relations
\[ b_i = b_i + x_{ki} \quad \text{followed by} \]
\[ a_{ij} = a_{ij} + \frac{1}{k-1} \left[ x_{ki} (kx_{kj} - b_j) + b_j \left( \frac{b_j}{k} - x_{kj} \right) \right] \]
for \( k = 2, \ldots, n \). Set \( m = n-1 \). If a regression intercept is not called for, set
\[ a_{ij} = b_i = 0 \quad 1 \leq i \leq j \leq p \]
and apply the recursion relations
\[ a_{ij} = x_{ki} x_{kj}, \quad b_i = b_i + x_{ki} \]
for \( k = 1, \ldots, n \). Set \( m = n \). Save the transgenerated data on tape.

2. Replace B and A by a mean vector and covariance matrix
\[ b_i = b_i/n \quad i = 1, \ldots, p \]
\[ a_{ij} = \frac{1}{m} a_{ij} \quad 1 \leq i \leq j \leq p \]
and compute the standard deviations
\[ s_i = \sqrt{a_{ii}} \quad i = 1, \ldots, p \]
If called for, print the means, standard deviations, and covariance matrix.

3. Replace A by a correlation matrix
\[ a_{ij} = a_{ij}/s_is_j \quad i, j = 1, \ldots, p \]
and print if called for.
4. Read a Sub-problem Card and restore the matrix $A$

$$a_{ij} = a_{ji} \quad 1 \leq j \leq i \leq p$$

$$a_{ii} = 1 \quad i = 1, \ldots, p$$

Set up the control vector $C$

$$c_i = \text{control value for the } i^{th} \text{ variable}$$

and set the number of variables in the equation $q = 0$.

5. Call subroutine STEPRG. This subroutine will step $A$, $C$, and $q$, and send back an index $k$ and a flag.

$$\text{Flag} = \begin{cases} 
-1 & \text{variable } k \text{ removed} \\
0 & \text{go to Step 9} \\
1 & \text{variable } k \text{ added}
\end{cases}$$

6. Let $d$ denote the index of the dependent variable. Compute and print:

- The residual degrees of freedom, sum of squares, and mean square.
  $$\text{df} = m-q, \quad \text{SS} = m s_d^2 a_{dd}, \quad \text{MS} = \text{SS}/\text{df}$$

- The regression degrees of freedom, sum of squares, mean square, and $F$ value.
  $$\text{rdf} = q, \quad \text{RSS} = s_d^2 - \text{SS}$$
  $$\text{RMS} = \text{RSS}/\text{df} \quad F = \text{RMS}/\text{MS}$$

- The standard error of estimate and the multiple correlation coefficient.
  $$S = \sqrt{\text{MS}}, \quad R = \sqrt{1 - a_{dd}}$$
7. A variable with index \( i \) is in the regression if \( c_i \leq 0 \). For each such variable, compute and print the regression coefficient, its standard error, and \( F \) value.

\[
\beta_i = \frac{s_i a_{id}}{s_i}, \quad S_i = \frac{S}{s_i} \sqrt{\frac{a_{ii}}{m}}, \quad F_i = \left( \frac{\beta_i}{S_i} \right)^2
\]

If the regression intercept \( \alpha \) is called for, compute and print

\[
\alpha = b_d - \sum_i \beta_i b_i
\]

where the summation is over all \( i \) such that \( c_i \leq 0 \).

8. A variable with index \( i \neq d \) is not in the regression if \( c_i \geq 1 \). For all such variables, compute and print the tolerance level, partial correlation coefficient, and \( F \) value.

\[
T_i = a_{ii}, \quad R_i = a_{id} / \sqrt{a_{ii} a_{dd}}
\]

\[
F_i = a_{i}^2 (m-q-1)/(a_{ii} a_{dd} - a_{id}^2)
\]

9. If called for, compute and print the residuals

\[
r_i = \alpha + x_{id} - \sum_j \beta_j x_{ij} \quad i = 1, \ldots, n
\]

where the summation is over all \( j \) such that \( c_j \leq 0 \). If a plot of residuals against the \( j \)th variable is called for, plot

\[
(r_i, x_{ij}) \quad i = 1, \ldots, n
\]

for each such \( j \).

10. Return to Step 4 and continue reading Sub-problem Cards until there are no more.

11. Return to Step 1 and continue reading Problem Cards until there are no more.
1. Compute

\[ V_{in} = \frac{F_{in} \cdot a_{dd}}{F_{in} + m - q - 1} + 2 \]

\[ V_{out} = \frac{F_{out} \cdot a_{dd}}{m - q - 7} \]

where \( F_{in} \) and \( F_{out} \) are the F values to enter and remove.

2. Find the \( k \) such that \( c_k < 1 \) and \( k \) minimizes

\[ V_k = c_k - \frac{a_{kd}}{a_{kk}} \]

If \( V_k < V_{out} \), set Flag = -1, increase \( c_k \) by 9, and go to Step 4.

If \( V_k \geq V_{out} \) or \( c_k \geq 1 \) for all \( k \), go to Step 3.

3. Let Tol be the minimum tolerance. Find the \( k \) such that

\( c_k > 1, a_{kk} \geq Tol, \) and \( k \) maximizes

\[ V_k = c_k + \frac{a_{kd}}{a_{kk}} \]

If \( V_k > V_{in} \), and \( m - q - 3 + a_k > 0 \), set Flag = 1, reduce \( c_k \) by 9, and go to Step 4.

If \( V_k < V_{in} \), and \( m - q - 3 + a_k \leq 0 \), or there is no \( k \) such that \( c_k > 1 \) and \( a_{kk} \geq Tol \), set Flag = 0 and go to Step 5.

4. Call subroutine STEP and return.

5. Set \( m = m + \text{Flag} \) and return.

Subroutine STEP

1. Set

\[ u_i = \begin{cases} 
    a_{ik} & 1 \leq i < k \\
    -\text{Flag} & i = k \\
    a_{ki} & k < i \leq p 
\end{cases} \]

\[ a_{ik} = a_{kj} = 0 \quad i \leq k < j \]

Revised: 9-1-65
2. Step the upper diagonal part of $A$

$$a_{ij} = a_{ij} - \frac{u_i u_j}{a_{kk}} \quad 1 \leq i \leq j \leq p+1$$

and return.

6. **REFERENCE**

SELECTED BIBLIOGRAPHY

A. BOOKS


B. ARTICLES


Lloyd, J. "Developing Creativity in the Culturally Deprived." The Instructor, 75 (February, 1966), 116.


C. REPORTS


D. UNPUBLISHED MATERIALS


