DOCUMENT RESUME

ED 048 318 24 TE 499 840

AUTHOR	Long, Newell H.
TITLE	Pilot Study for the Development of Music Discrimination Tests for Elementary School Children. Final Report.
INSTITUTION	Indiana Univ., Bloomington.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-9-0039
PUB DATE	Jun 70
GRANT	OEG-9-9-240039-0026
NOTE	77p.
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Elementary Education, *Listening Comprehension, Listening Skills, *Music, Music Appreciation, Music Education, Perception, *Perception Tests, *Test Construction, Testing

ABSTRACT

This pilot project proposed to construct a test or battery of tests to measure (1) the music discrimination abilities of school children aged 12 and under; (2) children's understanding of such basic concepts as rhythm, melody, and harmony; and (3) the ability of children to apply these concepts in music listening. Musical excerpts representing a balance among Baroque, Classical, Romantic, and Modern-Contemporary stylistic periods, which ranged in length from 6 to 30 seconds, were copied and "mutilated" to produce a pool of 243 possible test items. To compose mutilated versions, only one of the elements--rhythm, melody, or harmony--was changed, while a distortion was devised that was inherently weak by being erratic, inconsistent, illogical or mixed up. Items were recorded and evaluated on a five-point scale of difficulty for elementary pupils. Based on three out of four earlier experimental tests, two forms of the fifth test--composed of two parts--were administered (1 week apart) to the same 448 elementary school children. Evaluation of the test and recommendations for improving it were then made. [Data on this test, conclusions, a bibliography, sample answer forms, lists of participating schools and personnel, tables of results, and figures containing mutilations are provided. Not available in hard copy due to marginal legibility of original document. (JMC)]

FINAL REPORT
Project No. 9-0039
Grant No. OEG 5-9-240039-0026

PILOT STUDY FOR THE DEVELOPMENT OF MUSIC DISCRIMINATION TESTS FOR ELEMENTARY SCHOOL CHILDREN

Newell H. Long

Indiana University

Bloomington, Indiana

June 1970

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

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

Office of Education
Bureau of Research

Arts and Humanities Division

ACKNOWLEDGMENTS

The project has benefitted from the advice and opinions of Dr. Marilyn Pflederer Zimmerman, Dr. Kate Hevner Mueller, Dr. Clinton Chase, Dr. Richard Pugh and members of my graduate classes at Indiana University. Grateful acknowledgment of their assistance is hereby extended.

For their expert preparation and direction of musicians in the recording sessions thanks are due Prof. David Dawson and Prof. Henry Gulick and to Prof. Oswald Ragatz and Edwin Keenan as well for their skillful playing of the organ and piano items.

Special commendation is due Diana Polley who assisted me in various aspects of the test development, especially in administering the tests to children.

The cooperation of various Indiana school systems in permitting experimental testing of pupils and the help of both music teachers and elementary classroom teachers in making arrangements for the test sessions are deeply appreciated A list of the schools and cooperating personnel may be found in Appendix A.

TABLE OF CONTENTS

SUMMARY

The purpose of this study was to investigate the possibility of developing tests of music discrimination for elementary school children. The project resulted in the construction of such a test.

Assumptions made in pursuing the project were: (1) Music discrimination, while it requires perception of musical sounds, is the ability to make judgments about the artistic merit or "rightness" in the sounds and their organization. (2) A child's response to a piece of music may be made to the music in its entirety, to the totality of the musical elements in combination, whether or not there is awareness and response to the constituent parts. (3) An aesthetic response does not necessarily have to be an analytical one, but may approach a gestalt. (4) A test of music discrimination should present genuine music in full context as material about which musical judgments are to be made. (5) The musical materials for a music discrimination test should represent a variety of historical styles and performing media. (6) Children in primary grades (1 - 3) have not acquired sufficiently clear concepts of rhythm, harmony, or melody for them to be expected to identify the musical element that has been altered to produce a less acceptable version of a piece of music.

The test development began with the selection of 159 pieces of Western concert-type music and the arrangement of one or more mutilations of the first phrase of each until a total of 243 potential test items had been scored. The experience of the project director in his 1963-64 revision of Hevner's Oregon Music Discrimination Test helped both in the selection of music for items and in the writing of mutilations. The music represented in balanced amounts four style periods (Baroque, Classical, Romantic, and Modern-Contemporary) and the deformations were rhythmic, harmonic or melodic (no combinations). The mutilations were in themselves inconsistent and irregular so that judgments about the superiority of the original version over the mutilated one would not be a preference for one style over another, but about the artistic integrity and rightness of the original version over the illogical, ununified content of the deformation.

The test items and their mutilations, varying from five to thirty seconds in length, were recorded with high fidelity equipment on magnetic tape with performances by advanced university students or faculty musicians. Piano, organ, string quartet and woodwind quintet were the media used.

The recorded items were heard by four music faculty members and a graduate music education class to evaluate the quality of the music and its performance and to assign ratings of the probable difficulty of items for judgment by elementary school pupils.

Items rated easy or very easy were then selected for inclusion in five experimental tests: I, containing a mixture of items with respect to which element, if any, had been changed; II-5,

containing only items with rhythmic changes or no changes (identical versions); II-RR, containing no identical versions, but only those with rhythm altered; III-H, containing only harmonically altered items; and IV-M, containing only items with melodic mutilations. Each of these experimental tests, assembled by splicing tapes, had 19 or 20 "pairs of pieces" (phrases and their mutilations), of which three or four were designated practice items.

Each of the experimental tests was given to 100 or more pupils in grades two to five in test sessions of 20 minutes or less. Pupils were instructed to mark which version was better, the first or the second, or if they were the same. ("Same" items were discarded early in the project as too complicating.) Responses made on a very simple answer sheet were transferred afterward to scanning sheets for machine scoring and automatic preparation of punch cards needed in the item analysis by computer. The discriminating power of items and the reliability coefficients for these experimental tests varied widely from class to class and a great many items were too easy, having been responded to correctly by more than 75% of a class. Nevertheless a sufficient number of items survived the computer program's criteria for "good" test items for a "master test" - Test V - to be assembled.

This "master test" contained 58 items, 10 of which were practice examples. The test was divided into two parts which were given one week apart to a total of 383 pupils in five unlike communities. (The first half of the test was taken by an additional 65 pupils in a sixth community.) Statistical analysis of the results is shown in Table A.

TABLE A. MEAN SCORE, S. D., RELIABILITY, AND S. E. FOR TEST V

Grade	N	Mean	S.D.	Reliability by Kuder Richardson	S.E.	Reliability by Spearman-Brown	S.E.
1	67	16.3	3.6	.55	2.4	.68	2.0
2	99	17.4	2.9	.28	2.5	.44	2.2
3	68	17.4	3.0	.33	2.5	.49	2.1
4	21	19.4	2.8	.40	2.2	.63	1.7
5	128	20.2	3.2	.53	2.2	.55	2.1
All	383	18.2	3.5	.53	2.4	.59	2.2

Item analysis showed that the discriminating power of some items was less good than had been expected from their performance in the experimental tests. Ten items were too low in discriminating power or too easy and needed to be replaced with better ones. Forced choice design resulted in some guessing by pupils, but the irregularity of results from class to class suggests guessing was not greatly prevalent, for chance scores would have had an equalizing effect which was definitely missing.

The reliability of the test, which should improve when the weaker items have been replaced, is sufficiently high, nevertheless, to justify the use of the test, even in first grade, especially as a group measure.

A parallel, equivalent form should be developed to serve the needs for evaluation in music education research employing pre-test--post-test experimental designs.

Norms for both forms should then be sought.

The use of jazz, ethnic musics, and electronic music as test items should be explored.

INTRODUCTION

The Problem

The purpose of this pilot project was to construct a test or a battery of tests which would measure the music discrimination abilities of school children aged 12 and under and also measure children's understanding of basic concepts such as rhythm, melody, and harmony and the ability of children to apply these concepts in music listening.

Prior Studies

The success of music educators in developing musical skills in children is readily demonstrated in the quality of singing and instrument playing in classroom or public performance. The evaluation may be largely subjective, but it is not difficult. Measurement of cognitive learning in elementary school music has been largely through informal teacher-made tests, although there have been efforts at more precise and standardized measurement, e.g., Kwalwasser-Ruch "Test of Musical Accomplishment," Knuth "Achievement Tests in Music," and Colwell "Elementary Music Achievement Tests." However, these forays into achievement testing in music have been over-shadowed by the development of tests of music aptitude, for music aptitude tests have dominated the field of music testing since Carl Seashore first constructed his "Measures of Musical Talents" in 1919.

Quantitative assessment in the area of music appreciation was attempted in 1934 by Kate Hevner with her development of the Oregon Music Discrimination Test and there had been no other major effort in this field except for the revision of the Oregon Test into the Indiana-Oregon Music Discrimination Test in 1965 by the director of this project.

The Oregon Test consisted of 48 short excerpts from the beginnings of pieces of piano music together with a mutilated version of each excerpt in which one element, rhythm, melody or harmony, had been altered. Subjects responded by indicating which version they thought was the preferred (correct, unaltered) one and which element had been changed. The 1965 Indiana-Oregon revision on high fidelity magnetic tape included string quartet, woodwind quintet and organ items as well as piano ones, it extended the repertorial range to include more Baroque and Contemporary items, and it included items in which the two versions were identical.[1]

---

[1]Long, Newell H., "A revision of the University of Oregon Music Discrimination Test," (Ed.D. dissertation, Indiana University, 1965) 65-14,054.

In recent years several efforts to measure aesthetic responses in music have been reported, but most tests constructed for this purpose were designed for use with adolescents or adults. Crickmore in England developed a "syndrome test" wherein subjects responded on a 5-point scale (later reduced to a 3-point scale) to 22 recorded pieces of music representing a variety of styles.[2] A "Test for the Perception of thematic Relationships" was constructed by Haack for use in secondary schools.[3]

Student teachers were the subjects used by Lifton in his experimentation with a test of music reaction attempting to measure affective and aesthetic sensitivity.[4] To study the Leonard Meyer's concept of expectation as a principal factor in generating aesthetic responses, Simon and Wohlwill designed a test in which subjects responded to three 20-second versions each with "increasing amounts of variation and departure from expectation."[5] They used college undergraduates as subjects.

Both Petzold[6] and Pflederer[7] in designing ways of testing small children's perception in music relied on individual performance or oral verbalizing rather than written group testing, but their findings provided some guidance for the present study. Bridges asked children, K-3, to evaluate the appropriateness of different accompaniments to the same melodies and devised simple

---

[2]Crickmore, Leon, "An Approach to Measurement of Music Appreciation," Journal of Research in Music Education, Vol. XVI, No. 3, Fall 1968, pp. 239-255.

[3]Haack, Paul, "A Study of the Development of Music Listening Skills of Secondary School Students," (Dissertation, University of Wisconsin, 1966).

[4]Lifton, Walter M., "Development of a Music Reaction Test to Measure Affective and Aesthetic Sensitivity," Journal of Research in Music Education, Vol. IX, No. 2, Fall 1961, pp. 157-166.

[5]Simon, Carol R. and Wohlwill, Joachim F., "An Experimental Study of the Role of Expectation and Variation in Music," Journal of Research in Music Education, Vol. XVI, No. 3, Fall 1968, pp. 227-238.

[6]Petzold, Robert G., "Development of Auditory Perception of Musical Sounds by Children in the First Six Grades," Journal of Research in Music Education, Vol. XI, No. 1, Spring 1963, pp. 21-43.

[7]Pflederer, Marilyn and Sechrest, Lee, How Children Conceptually Organize Musical Sounds, Cooperative Research Project No. 5-0256, Northwestern University, 1968, 232 pages.

answer sheets for their responses to paired versions.[8]

Construction of a music discrimination test in Great Britain by Taylor probably provided the closest parallel to the project here reported.[9] Taylor composed twelve short pieces and four versions of each into which he introduced "errors of pitch, melody, rhythm and harmony." Piano, guitar, recorders, clarinet, bassoon and percussion were used in various combinations to record the test items. Scores upon administration of the test to more than 1,000 children showed an increase from age-group to age-group and reliable coefficients ranged from .76 to .9j. Taylor also devised a Music Preference Test of 30 items covering six stylistic periods; subjects expressed a preference when presented with paired examples from different periods. The 6-7 year group showed little common preference, seeming to like music from all periods equally well, but by age 8-9 tendencies were manifest: boys and 11-14 year olds showed strong preference for Period 5 (Brahms/Tschaikowsky); adults and girls expressed negative attitudes toward Period 6 (Stravinsky/Hindemith) and positive preferences toward Period 2 (Bach/Handel) and Period 3 (Haydn/Mozart).

Herbert Wing ostensibly developed his Tests of Musical Intelligence as aptitude tests, for he based their validity on their prediction of success in instrumental study and performance. Nevertheless he included paired performances of phrases of music, with one purposely spoiled either through intentional changes in rhythm, accent, harmonization, intensity and phrase grouping. These four sub-tests, which are all piano items presented on tape, are admittedly measures of tasteas well as perception. This approach is similar to that first employed by Kate Pevner in the Oregon Test. The ability to respond to Wing's items depends on acculturation and probably represents music discrimination as much as it does musical aptitude.

Edgar Gordon's Musical Aptitude Profile, an important new test of aptitude has a section for measuring "Musical Sensitivity" which likewise requires testees to select the better of two versions of a phrase of music heard from magnetic tape. Evidently Gordon considers discrimination (the ability to make correct judgments about music) a facet of music aptitude. However, he does not /state/ that the "Musical Sensitivity" sub-test is a discrimination test.

---

[8]Bridges, Virginia Ann, "An Exploratory Study of the Harmonic Discrimination Ability of Children in Kindergarten Through Grade Three in Two Selected Schools," (Ph.D. Dissertation, The Ohio State University, 1965, 65-13,206), reviewed by Kaplan, Barbara Connally, Council for Research in Music Education Bulletin, No. 19, Winter 1970, pp. 30-33.

[9]Taylor, Sam, "Development of Children Aged Seven to Eleven," Journal of Research in Music Education, Vol. XVII, No. 1, Spring 1969, pp. 100-107.

Need for the Study

The need for more and better tests in music was pointed out in a report of the Committee on Objectives and Evaluation to the Research Council of the Music Educators National Conference on March 11, 1968: "Much of the past effort to evaluate musical behavior and instruction has been ineffective because of failure to define the expected musical outcomes in operational terms. Further, this past effort, particularly as represented by teacher-made tests, has:

1. lacked profundity and rigor,
2. been badly fragmented,
3. failed to define objectives adequately,
4. frequently relied upon subjective judgment when objective measurement was needed, and
5. resulted in instruments devised and administered by persons lacking appropriate expertise.

"Evaluation instruments are needed which avoid the shortcomings cited and which permit an examination of the diverse musical behaviors characteristic of the American population. This urgent and growing need for effective evaluation has been compounded by:

1. the demand by federal agencies for precise evaluation of funded projects,
2. the problems of curriculum variance as related to population mobility,
3. the problems associated with the varying effectiveness of teachers,
4. widely differing attitudes among teachers and citizens concerning curriculum content, and
5. failure to recognize the diversity of possible musical behaviors."

This project for the construction of a new test or tests of music discrimination for elementary school children was, in part, proposed as a response to a memorandum from the Department of Health, Education and Welfare Office of Education, dated May 20, 1968, announcing an interest in "pilot studies directed toward the development of means of evaluating the non-cognitive aspects of education in the arts, particularly those in the affective domain." The memorandum further stated: "Art education [this includes music education, we assume] have been reluctant to evaluate their programs objectively, and tests of 'appreciation' have been largely cognitive and often subjective in nature. Interest in the evaluation of affective behavior has sharply increased as pressure on teachers for objective data concerning the effectiveness of their programs has grown. Some of the behaviors that have not yet been measured (affective) are crucial to the central purposes of arts education, while many of those that can easily be measured (cognitive) tend to be peripheral and superficial. The time is at hand for a concentrated attack on

the problem." Music discrimination, which this project seeks to measure, clearly falls within the affective domain.

When administering the Indiana-Oregon Music Discrimination Test to 4,400 students in 1967 to obtain norms, the author became keenly aware that, while that test worked fairly well with sixth grade pupils, it was too difficult and the reliability was low (0.38) with fifth graders.[10] An adaptation of the design of that test and utilization of musical materials expressly selected for a simpler test appeared desirable and feasible.

Assumptions and Guide Lines

Music discrimination, the ability to make judgments about music, is assumed to be a product of acculturation and to be influenced by the amount and nature of musical experiences people have had. It represents a kind of connoisseurship even though it may emerge at a very elementary level.

It is assumed that it is possible to measure music discrimination, an aspect of music appreciation, even though the latter has not been scientifically defined. Appreciation of music is not a unitary capacity, but a complex of different sensitivities which will be possessed to varying degrees by different individuals. There are various levels and intensities of appreciation rather than discrete, opposite conditions of awareness and unawareness of the quality and components of music.[11]

A discriminative response to music, if it is to represent a valid musical judgment, must be made to music which is complete, having its various elements present and sounding together in normal, performance context. The musical person must be able to keep distinct his perception of the interwoven parts, hearing them in the context of the composition, while also attending to the total tonal effect.[12]

---

[10]Long, Newell H., "Establishment of Standards for the Indiana-Oregon Music Discrimination Test Based on a Cross-Section of Elementary and Secondary Students with Analysis of Elements of Environment, Intelligence and Musical Experience and Training in Relation to Music Discrimination." U.S. Office of Education Grant No. OEG-3-7-700027-2893, Indiana University, 1969, p. 1.

[11]Hevner, Kate, "Appreciation of Music and Tests for Appreciation of Music," in Studies in College Teaching, Vol. 4, No. 6, pp. 83-151.

[12]Révész, G., Introduction to the Psychology of Music.

A response to music need not be analytical or atomistic. The author found in studying the test responses made by children in grades 5 through 9 to the Indiana-Oregon Test that scores on the A-B-No portion of the test ran consistently higher than on the R-H-M portion, indicating that students must have reacted to a gestalt, correctly making a judgment about the music in its entirety more often than they did about its elements. This view is supported by Pike who claims, "The structure of immediate musical experience is obtained through intuitive cognition, a direct and habitual perceptual penetration in which the tonal event and its meaning are grasped at the level of its essential structure. . . . Particular tonal events do not make the work of art. The experience of the various tonal gestalten adds up to the total experience of the composition as the tonal gestalten unfold their multiple qualities during the process."[13]

Crickmore, commenting on the nature of music appreciation, says enjoying music seems to involve both an active and a passive element. "In moments of profoundest involvement the enjoyment of music is felt as a kind of effortless awareness, more passive or receptive than active--an intuitive act which involves no discursive or reflexive process."[14] Citing Wing as authority he points out that most tests of music appreciation involve critical comparisons or choices between two or more music items. In his own test construction Crickmore assumed that music may be listened to in many ways, one of which, with "insight," he claims is especially compatible with the nature of aesthetic perception.

Sensitivity to style in music, an aspect of appreciation which has been receiving increasing emphasis, is allied with and dependent upon sensitivity to those elements which determine style: rhythm, harmony, melody texture and dynamics. With the possible exception of form the most significant elements in music are rhythm, melody and harmony; consequently, it is pertinent to attempt to measure discrimination with emphasis on mutations of these elements. Copland warns us that we are seldom conscious of hearing these elements separately, but are usually concerned with their combined effect.[15]

It is impractical to include musical form as an element for testing because pieces would have to be heard in their entirety before judgments about form could be made. Not only would the

---

[13]Pike, Alfred, "The Phenomenological Analysis and Description Experience," Journal of Research in Music Education, Vol. XV, No. 4, Winter 1967, pp. 316-319.

[14]Crickmore, op. cit., pp. 239-240.

[15]Copland, Aaron, What to Listen for in Music.

playing of entire pieces be too time consuming, but the degree of musical memory involved would be considerable. However, the ability to perceive musical form is dependent upon abilities to discriminate concerning the elements (rhythm, harmony and melody); therefore, measurement of discriminations when these elements are varied should yield some, partial measurement of ability to appreciate form.

While, as Mursell claims,[16] musical sensitivity includes a feeling for emotional and expressive values in the music, it is assumed that the aptness or fitness of a musical phrase to convey its musical meaning may be judged without actually identifying or analyzing the emotion or feeling.

Appreciation of music does not mean the same thing as enjoyment of music, although these two behaviors may be concurrent and may interact. It should be possible, then, for a person to make valid judgments about music he neither likes nor enjoys.[17]

To insure the artistic validity of the evolving test only music by composers of repute is used. Furthermore, it is assumed that the artist composer, skilled in his craft, has arrived at superior aesthetic solutions to his creative problems; consequently, mutilated versions of his work, even phrases from it, will be less satisfactory aesthetically than the composer's original version.[18] The author found, when experimenting with items for the Indiana-Oregon Music Discrimination Test, that he could arrange lush modern harmonizations of classical items which experienced musicians would prefer, that is, rate artistically higher, than the originals, but he also learned that it was not difficult to "compose" mutilations which would avoid this hazard.

Neither the Oregon Test nor its revision, the Indiana-Oregon Test, nor the test resulting from this project is considered a comprehensive measure of the ability to appreciate music or of aesthetic sensitivity. Rather they are considered instruments for evaluating certain aspects of listening and judgment which function in making musical discriminations.

All we can conveniently test is the ability of subjects to deal with some of the listening aspects for a short period--the ten or twenty seconds of a complete phrase, in fact. We shall assume

---

[16] Mursell, James L., "Applications of Psychology to the Arts," Teachers' College Record, 37:290-299, Jan. 1936.

[17] Hevner, op. cit., p. 108.

[18] Ibid. p. 112.

tentatively that the ability to deal thus with phrase units is symptomatic of the listener's ability to deal with the entire work of art.

While sensual enjoyment might result from hearing various tone colors and satisfaction might accrue to the listener from successful identification of instruments, discriminations concerning timbre, as an element of music, were not included in the test because it was assumed that perception of timbre is auxilliary to, rather than fundamental to, understanding the music. Admittedly this is less true with respect to the music of Webern and today's avant garde composers, but for music of the "common practice" periods, the assumption seems justified. Furthermore, Petzold has reported that his study had found item difficulty to be a function of melodic content, but not of timbre.[19]

Nearly all test items were selected from that music literature which is sometimes called "serious music" or, as Sigmund Spaeth put it, "permanent music." Art music, rather than folk or popular music, provided the phrases for test items, because a generally accepted goal of music educators is the development of ability to attend to art music and to recognize excellence or the lack of it therein. This limitation does not imply that aesthetic responses may not be induced by popular or folk music; instead, an assumption was made that because art music is associated with high levels of aesthetic response, the test should depend upon responses to music of accepted artistic merit. This assumption does not disregard Greenberg's observation that the influence of constant exposure to popular music on children's feelings for tonality and cadence in music of all types should be recognized.[20]

Two familiar pieces, "America" and Jingle Bells," were included in the item pool for use in practice items or for the purpose of giving children confidence in their responses.

Discriminations concerning quality of performance are not included in the test, although these might represent one ability of the musical connoisseur. Recognition of superior piano or violin playing or the identification of faulty rendition of songs or sonatas may indeed bring moments of satisfaction to listeners, but such critical powers, so dependent upon extended listening experience

---

[19]Petzold, Robert G., "Auditory Perception by Children," Journal of Research in Music Education, Vol. XVII, No. 1, Spring 1969, p. 85.

[20]Greenberg, Marvin, in review of Peterson, Agda Viola, "A Study of Developmental Listening Factors in Children's Ability to Understand Melody," dissertation, University of Rochester Eastman School of Music, 1965, in Council for Research in Music Education Bulletin, No. 16, 1968(?).

and specific knowledge of instrumental or vocal technique, we consider peripheral to discriminations concerning music content, the kind of discriminations we felt obliged to try to measure.

Measurement of attitudes toward music were likewise omitted, even though the development of favorable attitudes is frequently listed as a goal of music instruction. To measure the child's awareness of the quality of the music itself was assumed to be a sufficiently challenging goal.

PROCEDURE

Development of Test Items

After reviewing the literature and research reports relevant to the project, the author set about selecting phrases of music that might be used in a test of discrimination. The criteria employed were:

1. Phrases must be from the music of composers of recognized stature and mastry.
2. Phrases must be the first phrases in pieces or movements so that they will not be heard out of context.
3. The items should be relatively unfamiliar.[21]
4. The items should represent a reasonable balance between Baroque, Classical, Romantic, and Modern-Conterporary stylistic periods.
5. There should be a variety of tempi, and modes represented in the items.
6. The media of piano, organ, string quartet and woodwind quintet should be about equally represented.
7. There should be interr balance in the application of criteria 4, 5 and 6.

In all 159 musical excerpts ranging from six to thirty seconds in length were copied and one or more mutilations of each "composed" until a pool of 243 possible items had been accumulated. The count of potential test items in each category is presented in Table 1.

In composing (or decomposing) the mutilated versions great care was taken to change only one of the elements--rhythm, melody or harmony--in any one mutilation. This was done to prevent ambiguity should testees be required to identify which element had been altered. The degree of deviation from the original versions was varied so as to provide subtle (difficult) items, obvious (easy) ones, and others in between.

Since the basic design of the test expected the subject to choose which was the better version, the original or the mutilated one, it was not enough to just change the rhythm, melody or harmony, because it was necessary to devise a distortion that was inherently weak by being erratic, inconsistent, illogical or mixed up. For example a Mozart item reharmonized to achieve a less artistic version might have measures of traditional harmony randomly interspersed with measures of contemporary harmony, chords of fourths, for example. Typical items and their mutilations are presented in notation along with item analysis in the next chapter.

---

[21] Lifton, Walter M., op. cit. p. 159.

Table 1. NUMBER OF ITEMS IN RECORDED POOL BY HISTORICAL STYLE PERIOD AND PERFORMANCE MEDIUM

Medium	Ba-roque	Classi-cal	Roman-tic	Modern-Con-temporary	Total
Piano Excerpts		13	14	12	46
Changed Rhythmically	3	5	9	3	20
Changed Harmonically	3	6	6	5	20
Changed Melodically	3	5	4	8	20
Tempo: Fast	4	11	8	4	27
Medium	2	3	9	8	22
Slow	3	2	2	4	11
Additional shorter versions	2	2	0	1	5
Added melody-only versions	0	2	0	0	2
Total potential items	11	20	19	17	67
Organ Excerpts	14	3	20	8	45
Changed Rhythmically	2	3	14	2	21
Changed Harmonically	7	2	9	2	20
Changed Melodically	7	3	8	7	25
Tempo: Fast	1	0	10	2	13
Medium	7	3	8	7	25
Slow	8	1	4	1	14
Additional shorter versions	3	0	3	1	7
Added melody-only versions	0	1	1	0	2
Total potential items	19	9	35	12	75
String Quartet Excerpts	7	10	12	7	36
Changed Rhythmically	3	4	4	5	16
Changed Harmonically	3	2	8	1	14
Changed Melodically	2	6	5	2	15
Tempo: Fast	1	2	2	4	9
Medium	4	4	6	3	17
Slow	3	6	9	1	19
Additional shorter versions	0	1	1	0	2
Added melody-only versions	0	1	1	0	2
Total potential items	8	14	19	8	49
Woodwind Quintet Excerpts	7	9	2	14	32
Changed Rhythmically	4	5	0	6	15
Changed Harmonically	2	4	0	6	12
Changed Melodically	5	5	2	7	19
Tempo: Fast	0	4	0	4	8
Medium	4	9	1	12	26
Slow	7	1	1	3	12
Additional shorter versions	0	1	0	2	3
Added melody-only versions	0	2	0	1	3
Total potential items	11	17	2	22	52
Total number of excerpts mutilated and recorded					159*
Total number of potential test items recorded					243*

* These figures do not include 18 other excerpts and 25 deformations of them which were prepared, but not recorded.

Recording Sessions

Advanced student performers from Indiana University School of Music were employed to record the items and artist faculty members were employed to coach the ensembles prior to and during the taping sessions. Repeated "takes" were secured to make certain convincing, musical performances had been recorded. The Audio-Visual Department's experienced crew of tape recording experts operated and monitored the professional, high-fidelity equipment.

Most of the recording sessions were in Recital Hall late at night when there was minimum probability of interruption or infiltration of outside noises.

Two copies were dubbed from each master tape so that two identical versions of any item would be available.

Face Evaluation of Items

The director, his assistant, and the principal musicians involved with the recording sessions heard the recorded test items and evaluated them for both musical value and performance quality. Not only were "takes" of unsatisfactory quality identified, but some entire items are disqualified from further consideration.

A music education seminar also heard the tapes and rated the items on a 5-point scale as to anticipated difficulty for elementary school pupils. This seminar of 9 graduate students also identified items which they thought were artistically weak.

Experimental Test I

Guided by the difficulty ratings assigned by the seminar as well as by their own ratings, the director and his assistant selected nineteen items for the first experimental test. An even distribution of items between historical periods and between performing media was sought, but choices were confined to items rated "easy." Four of the items were picked to be used as practice items. All the items were cut from their respective tapes and spliced together to form Experimental Test 1. Whether the correct version or the mutilated one came first on the spliced tape was determined by the toss of a coin.

Experimental Test I contained one practice item and two other items in which the two versions were identical. The response sheet provided a column of 1's, a column of 2's and a column of the word "same." The instructions, which were given "live" by the administrator of the test, said to make a cross through the 1 if the first version sounded better than the second, through the 2 if the second version sounded better than the first, or through the word "same" if both versions sounded the same.

It will be noted here that the author's examination of other research has led him to believe that the concepts of rhythm, melody and harmony were not sufficiently formed in the lower grades for pupils to be expected to identify which element had been changed in a mutilation. Pflederer[22] found subjects across her age groups, 5 through 13, lacked vocabulary to discuss the music. Examination of the recorded oral responses of children in performing the various tasks in Pflederer's experiment showed that children in the lower elementary grades did not use the terms rhythm, melody or harmony. While these children may have formed, loosely or otherwise, these concepts, they could not be expected to identify them reliably by name. An example of the vagueness of their concepts is that 5-, 7- and 9-year olds thought the "tune" was different when the same melo · was presented by different instruments.

Andrews and Deihl[23] found that their fourth grade subjects confused the terms high, low, slow and fast. They observed that some children appeared to have concepts of pitch, loudness and duration, but confused the labels.

While Bridges [24] found that children in kindergarten through third grade could discriminate harmonically in varying degrees, her study did not require pupils to identify harmony as harmony.

(Later, in working with form V-MX of the test, the inability of second and third grade pupils to identify, during the practice items, which element had been changed further confirmed our view about the inadequacy of these concepts, rhythm, harmony and melody, at that level.)

Consequently, the responses expected in Experimental Test I were similar to those expected in the A-B-C portion of the Indiana-Oregon Test, but any expectation that pupils would identify the element changed to create a deformation of a musical phrase was abandoned. It was hoped that this simplification of response would help bring the music discrimination test to a lower-elementary level.

---

[22]Pflederer, Marilyn, How Children Conceptually Organize Sounds, Cooperative Project No. 5-0256, Northwestern University, 1965, 232 pages.

[23]Andrews, Frances M., and Deihl, Ned C., "Development of a Technique for Identifying Elementary School Children's Musical Concepts," Council for Research in Music Education Bulletin, No. 13, Spring 1968, pp. 1-7.

[24]Bridges, op. cit.

Trial Runs for Experimental Test I

Experimental Test I, with a mixture of rhythm-changed, harmony-changed, melody-changed and the-same items, was given to 55 third grade pupils in the Monroe County Community Schools.

The responses made by the children were transferred with clerical help to scanning sheets for machine scoring and computer analysis. This practice was followed in all subsequent testing.

The mean score on 15 items was 5.84 which is just barely above what one could expect by chance with a 3-choice response pattern. The reliability coefficients obtained were disasterously low: .02 by Kuder Richardson formula, -.04 by Spearman Brown. Nevertheless, the computer program labelled seven items as "good" in both discrimination and difficulty.

Before receiving this analysis we had already given the test to 39 seventh and eighth grade students at University School. We had originally planned to use this group in checking external validity. The mean score for this group was 7.36 with S.D. 1.8, but the reliability was even lower: Kuder Richardson -.11; Spearman-Brown -.06. Only four items were labelled "good."

Experimental Tests II-R and II-RR

Items were selected and spliced into a second experimental test before the scoring and item analysis of Test I had been completed. This test, labelled Test II-R, contained fourteen items in which rhythm had been changed to achieve the mutilation and four in which the two versions were identical. Three of the items were selected as practice items.

Test II-R was given to 66 third grade pupils in the Ellettsville Elementary School. The mean number correct on the fifteen items scored was 6.3 and the reliability computed by split-halves method was .24, which, corrected by the Spearman-Brown formula, was .39 for the entire test.

While Test II-R seemed to operate much better with third grade pupils than did Test I, it was decided to try a simplification of the response pattern by dropping the idea of items with identical halves. Accordingly the "same" items were replaced with four items which had rhythmic deformations. The II-R test thus modified was renamed Test II-RR.

Test II-RR was administered to five classes in the Wayne Township School in Bartholomew County with the results which are shown in Table 2.

TABLE 2. RESULTS FROM TEST II-RR AT WAYNE ELEMENTARY SCHOOL

Grade	2	3	3	4	5
Number Tested	28	34	33	28	30
Mean Score	8.04	8.26	9.33	9.93	9.13
Standard Deviation	2.15	2.18	1.96	2.32	2.42
Reliability					
Kuder-Richardson	.25	.29	.16	.43	.51
Standard Error	1.86	1.83	1.80	1.75	1.69
Reliability					
Spearman-Brown	.35	.70	.28	.48	.65
Standard Error	1.74	1.20	1.66	1.68	1.43
Number of items marked "good" by computer	1	3	6	2	9

Four items did not earn a "good" rating for any of the five classes; three items earned a "good" for one of the classes; five items earned a "good" in two classes; one item for three classes; and one item for four classes. (A "good" rating is given when the discriminating index is large enough, depending on the number of subjects, and the difficulty of the item is neither above 25% nor below 75%.)

Test III-H

Meanwhile a test composed entirely of items with harmonically mutilated versions was selected and assembled. Three practice items and fourteen test items were used.

This test, III-H, was tried out in a third grade room at Elm Heights School, Bloomington, and in one second grade, one third grade and one fourth grade in Templeton School, Bloomington. Table 3 gives the results.

TABLE 3. RESULTS FROM TEST III-H AT ELM HEIGHTS AND TEMPLETON

Grade	2	3	3*	4
Number Tested	20	20	31	20
Mean Score	9.35	10.25	11.84	10.20
Standard Deviation	2.94	2.12	1.97	2.73
Reliability				
Kuder-Richardson	.70	.46	.59	.73
Standard Error	1.62	1.56	1.25	1.41
Reliability				
Spearman-Brown	.75	.73	.51	.70
Standard Error	1.46	1.11	1.38	1.50
Number of items marked "good" by Computer	8	3	2	5

*Elm Heights School, other rooms at Templeton School (deprived area)

The number of items in which no class earned a "good" rating for it, 1; items with one class marked "good," 9; two classes, 2; three classes, 1; and four classes, 1.

The improvement in reliability of Test III-H over that for Test II-R encouraged the director to proceed with the formation of Test IV-M.

Experimental Test IV-M

Seventeen items, including three for practice, in which alterations of melody constituted the mutilation were assembled as Experimental Test IV-M. This was administered to five classes in Poston Road School, Martinsville, Indiana. Table 4 shows the results obtained.

TABLE 4. RESULTS FROM TEST IV-M AT MARTINSVILLE.

Grade	2	3	3	4	5
Number Tested	24	30	26	23	27
Mean Score	8.13	8.27	8.81	8.43	9.26
Standard Deviation	2.15	1.66	2.30	1.93	1.70
Reliability					
Kuder-Richardson	.36	-.06	.50	.22	.07
Standard Error	1.73	1.71	1.62	1.71	1.64
Reliability					
Spearman-Brown	.54	.35	.53	.38	.09
Standard Error	1.47	1.34	1.57	1.52	1.62
Number of items marked "good" by computer	5	3	4	5	2

The number of items in which its efficacy was rated "good" for no class was 4; for one class, 6; for two classes, 2; for four classes, 1; for all five classes, 1.

Up to this point the item analysis had revealed alarming inconsistency between classes and between grades in the discriminating power of the various items in various experimental tests. While it was apparent that considerable guessing instead of critical listening might be operating, speculation suggested that chance factors like guessing ought to lessen rather than increase the variation in responses between classes and grades. Speculation also suggested that these inconsistencies as well as the irregularities in the reliability coefficients were due to the fact that computations had all been made on small (one classroom) populations.

Furthermore, it was now discovered that the item-analysis by computer had not ignored the practice items, but treated them as omitted items. This had had a depressing effect on both reliability and discrimination indices.

Consequently, the project director decided to combine the most discriminating items from the experimental test into one longer test with the expectation that more consistency in discriminating power of items and greater reliability would emerge.

Test V

From experimental Test II-RR, Test III-H and Test IV-M thirty-three items were selected on the basis of discriminating power and suitable difficulty. A few items statistically too easy for a "good" test were included as practice items or to motivate continued listening through building up the listener's confidence.

Observations in administering the experimental tests had shown that twenty minutes was the maximum length of time pupils in the lower grades could be expected to listen attentively. Taylor[25] had estimated 25 minutes as the maximum time permissable for a music test battery for 7- and 8-year olds. He found a break within this period necessary.

To obtain a test long enough to insure reasonable reliability and keep within the intensive listening time span of young children, Test V was constructed in two parts, with the second portion to be administered a day or more after the first part.

The first part, designated V-RH, contained twenty items of which the first ten had mutilation through change of rhythm; the others had change of harmony. Three rhythm items and two harmony ones were practice items. The second part, designated V-MX, contained nine melody items at the beginning and then nine assorted ones. Two melody items were for practice as were three of the assorted ones. Each part required twenty minutes or less to administer.

Test V-RH was administered to three rooms of first graders (two schools and communities represented), four rooms of second graders (four schools and four communities), two of third graders (two schools and two communities), three of fourth graders (two schools and two communities), and four rooms of fifth graders (three schools, three communities). One week later Test V-MX was given to these same classes, with the exception of two fourth grade classes which could not be retested due to conflict of closing of school activities.

The responses from the two parts of the test, V-RH and V-MX, were transferred from the two answer sheets each child had used to one scanning sheet for machine grading. The single answer sheets

---

[25] Taylor, Sam, "Development of Children Aged Seven to Eleven," Journal of Research in Music Education, Vol. XVII, No. 1, Spring 1969, p. 101.

for pupils who were absent on one of the two testing days were discarded. Care was taken to omit the practice items; that is, practice items numbered 1, 2 and 3 were not recorded, of course, on the scanning sheet, but the response to item numbered 4 on the child's answer sheet was recorded as item No. 1 on the scanning sheet, etc.

Item analysis was done by computer in the Indiana University Research and Computing Center using the program normally employed by the School of Education Bureau of Educational Studies and Testing. Results of this item analysis and other data on Test V are shown in the next chapter.

Instructions to Pupils

In giving instructions to the pupils the word "test" was avoided. By using the analogy of taking a bit out of one apple and then another apple to see which one tasted better, the children were told to "taste" the music and decide which was the better of the two versions of each piece. For the two or three practice items on each section of the test the response pattern was placed on the chalkboard and the correct response demonstrated visually.

At the beginning of our testing we asked the children to indicate their answers by placing a cross through the numeral "1" or "2" or the word "Same", but at the suggestion of classroom teachers the children were asked to make a circle around "the right answer" instead. It was observed that children in the lower grades drew large circles, ovals and other shapes, but this did not appear to affect their level of concentration and it created no problem for the person who transfered the responses to the scanning sheets for machine scoring.

DATA ON TEST V

The pupils who were given form V of the music discrimination test during April and May 1970 were from six elementary schools.

A. Skiles Test School located in a high income, upper middle class suburb of Indianapolis
B. University School on the campus of Indiana University but operating as a part of Monroe County Community Schools and serving not only a part of the University community but also a low income rural area
C. Thornton School located in a middle income section of Terre Haute and
   Rea School located in a low income section near downtown Terre Haute (Results from Thornton and Rea Schools are combined in tables.)
D. Edgewood School in a lower middle class suburb of Indianapolis
E. Eastern School, a rural school in an agriculturally poor part of Greene County.

The responses of 448 children who took form V of the test are tabulated and analyzed below.

Table 5 shows by grades in these schools the number tested mean score and standard deviation. Also shown are reliability and standard error of measurement as computed by the Kinder-Richardson method and by the Spearman-Brown method.

It should be reported that less than ideal conditions existed for some of the testing, although all school personnel had been very cooperative. For example, it was a hot, humid day when one group of second graders was being tested, so the windows and an outside door were open for ventilation. In the middle of the testing session children from another classroom noisily began their recess activities on the playground area immediately adjacent to the classroom. A few minutes later a truck drove near the classroom windows to make a delivery. Although testing was halted until some of the disturbance had abated, the interruptions must have had a disturbing, if not depressing effect upon the test results.

TABLE 5. MEAN SCORES, S.D., RELIABILITY AND S.E. FOR TEST V

Grade	School	N	Mean	S.D.	Reliability Kudor Richardson	S.E.	Reliability Spearman Brown	S.E.
1	A	42	16.6	2.8	.26	2.4	.51	2.0
1	B	25	16.3	4.6	.74	2.3	.77	2.2
1	A+B	67	16.3	3.6	.55	2.4	.63	2.0
2	A	29	18.3	3.1	.40	2.4	.52	2.1
2	B	26	17.7	2.7	.22	2.4	.10	2.5
2	C	17	17.4	2.9	.38	2.3	.50	2.0
2	E	27	16.2	2.5	-.05	2.5	.25	2.1
2	A+B+C+E	99	17.4	2.9	.28	2.5	.44	2.2
3	C	44	17.8	2.9	.31	2.4	.54	2.0
3	E	24	16.6	3.4	.50	2.4	.72	1.8
3	C+E	68	17.4	3.0	.33	2.5	.49	2.1
4	B	21	19.4#	2.8#	.40#	2.2#	.63#	1.7#
4	D	65	12.3#	1.9#	.49#	1.4#	.50#	1.4#
5	A	58	20.6	2.9	.46	2.1	.47	2.1
5	C	30	19.5	3.0	.45	2.2	.50	2.1
5	E	40	19.7	4.1	.73	2.1	.58	2.7
5	A+C+E	128	20.2	3.2	.53	2.2	.55	2.1
All	A+B+C+E	383	18.2	3.5	.53	2.4	.59	2.2

#Only 1st half of test taken.

An item analysis by grade and school is reported on pages 25 to 59. An asterisk after a discrimination index indicates that this item met the internal validity criteria of the computer program employed by the Indiana University Bureau of Educational Studies and Testing. That program indicated acceptable items for the population tested, but rejected items if more than 75% marked the item correctly or if fewer than 25% marked the item correctly even though the discrimination index might be sufficiently high otherwise.

The analysis of the item is followed by a copy of the music used in each instances. These music scores show how representative mutilations were achieved and what kind of music produced

the greatest and least discrimination power through mutilation of each element--rhythm, melody or harmony--in this test.

Data for grade four is incomplete because school D was unable to schedule the second testing session. Consequently the statistics listed in the "all" catagory in Table 5 and at the bottom of the item analysis sheets do not include school D's fourth grade results.

The performance time given for each item includes the playing time of both versions and the three or four second pause between versions.

Composer Max Reger — Item No. for Test V - 1

Title Scherzo No. 5 — Recording Label
1st version P 20 A
2nd version P 20 C

Playing time 21 sec. Key C — Tempo Fast

Element Changed Rhythm — Medium Piano

Nature of Mutilation : Alternate (even numbered) measures of the 3/8 meter are replaced with 3/4, 2/4 and 5/8 measures. With the corresponding increases in time values for notes and/or rests the music is thus stretched out.

Used in Experimental Test II-RR as Item No. 16

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.39	75
3rd grade		.40 .28	65 70
4th grade		.51*	75
5th grade		.47*	73
In Test V			
	School		
1st grade	A	-.02	29
	B	.42*	32
	A+B		30
2nd grade	A	.16	52
	B	.06	81
	C	-.06	18
	E	-.10	74
	A+B+C+E		59
3rd grade	C	-.14	66
	E	-.09	75
	C+E		69
4th grade	B	.00	100
	D	.62	85
	B+D		90
5th grade	A	.12	98
	C	.31	83
	E	.67	90
	A+C+E		92
All grades	All	.28*	69

Composer Franz Danzi

Item No. for Test V - 2

Title Gypsy Dance

Recording Label
1st version W 21 A
2nd version W 21 B

Playing time 19 sec. Key g minor Tempo Fast

Element Changed Rhythm Medium Woodwind Quintet

Nature of Mutilation: A regular continuing succession of eighth-note chords comprising the accompaniment is interrupted by variously placed rests and 16th note chords. The melody in the flute is kept the same.

Used in Experimental Test II-RR as Item No. 14

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.58*	62
3rd grade		.62* .28	62 79
4th grade		.46*	64
5th grade		.35*	60
In Test V			
	School		
1st grade	A	.31*	62
	B	.08	46
	A+B		64
2nd grade	A	.33	69
	B	.35	46
	C	.64	76
	E	.24	33
	A+B+C+E		54
3rd grade	C	.33*	45
	E	.38	38
	C+E		43
4th grade	B	.68*	48
	D	-.04	54
	B+D		52
5th grade	A	.20	40
	C	.32	37
	E	.30*	50
	A+C+E		43
All grades	All	.23	50

W 2: A Gypsy Dance Fr. Dance

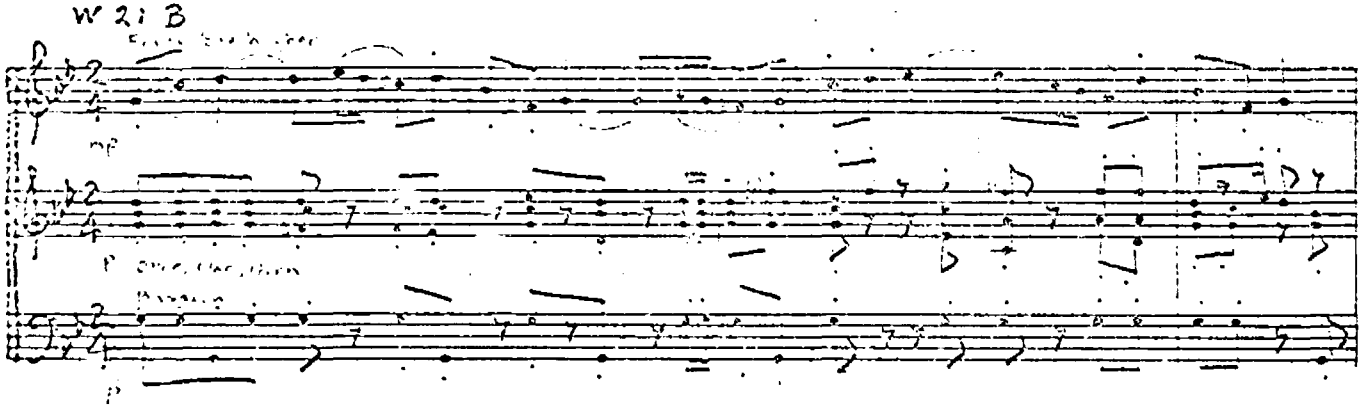


Figure 1. Score and Mutilation for Item 2. An example of rhythmic alteration to the accompaniment only. This is an item with low discriminating power.

Composer J. S. Bach

Item No. for Test V - 3

Title Cantata No. 167

Recording Label
1st version S 31 B
2nd version S 31 A

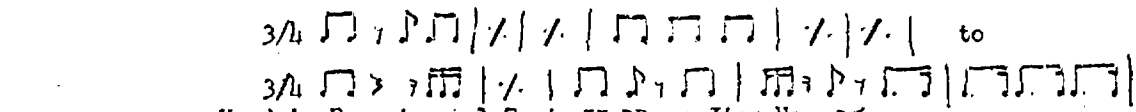
Playing time 27 sec. Key G

Tempo Medium

Element Changed Rhythm

Medium String Quartet

Nature of Mutilation: The rhythmic pattern of the three accompanying parts below the running 16th note melody are changed from:



to

Used in Experimental Test II-RR as Item No. 15

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.35	57
3rd grade	.32	.28	76 82
4th grade		.36	64
5th grade		.44	80
In Test V			
	School		
1st grade	A	.40	74
	B	.16	80
	A+B		76
2nd grade	A	.19	72
	B	-.06	73
	C	.47*	65
	E	.45*	56
	A+B+C+E		66
3rd grade	C	.40	77
	E	.59*	63
	C+E		72
4th grade	B	.18	81
	D	.43	80
	B+D		80
5th grade	A	.32	66
	C	.22	93
	E	.50	85
	A+C+E		87
All grades	All	.37	77

Composer Baird — Item No. for Test 7 - 4

Title Marcia from Divertimento — Recording Label: 1st version W 19 A; 2nd version W 19 C

Playing time 27 sec. Key atonal Tempo Fast

Element Changed Rhythm — Medium Woodwind Quintet

Nature of Mutilation: The notation is kept in 2/4, but rhythms are shifted so that eight measures become ten as some rhythmic units in the various parts are contracted and others enlarged with the contractions and enlargements unpatterned.

Used in Experimental Test II-RR as Item No. 5

		Discrimination Index		Difficulty Percent responding correctly	
In Experimental Test					
2nd grade			.37		36
3rd grade		.42*	.50*	35	55
4th grade			.36		64
5th grade			.30*		67
In Test V					
	School				
1st grade	A		.45		81
	B		.50		84
	A+B				82
2nd grade	A		.16		76
	B		.30		62
	C		.59		82
	E		.02		59
	A+B+C+E				69
3rd grade	C		.19		84
	E		.50*		75
	C+E				81
4th grade	B		.10		95
	D		.00		94
	B+D				94
5th grade	A		.15		95
	C		.22		97
	E		.66		85
	A+C+E				92
All grades	All		.38		83

Composer E. Batiste

Item No. for Test V - 5

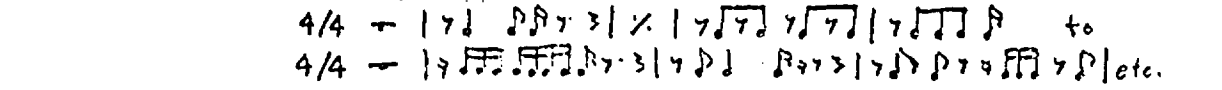
Title Offertory (St. Cecilia), Op. 9

Recording Label
1st version Q 9 B
2nd version Q 9 A

Playing time 27 sec. Key f minor Tempo Medium

Element Changed Rhythm Medium Organ

Nature of Mutilation : Rhythm of left hand chords made less regular. Changed from:



Used in Experimental Test II-RR as Item No. 6

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.40	36
3rd grade		.57* .59*	38 55
4th grade		.20	
5th grade		.33	
In Test V			
	School		
1st grade	A	.58*	77
	B	.44*	
	A+B		
2nd grade	A	.19	
	B	.06	
	C	108	
	E	.41	
	A+B+C+E		
3rd grade	C	.05	
	E	.28	
	C+E		
4th grade	B	.34	
	D	.45*	
	B+D		52
5th grade	A	.37*	66
	C	-.08	67
	E	.28	72
	A+C+E		69
All grades	All	.27*	60

Composer Alberto Ginastero

Item No. for Test V - 6

Title Sonata

Recording Label
1st version P 42 A
2nd version P 42 B

Playing time 33 sec. Key Obscure (C?) Tempo Fast

Element Changed Rhythm Medium Piano

Nature of Mutilation: Meter and/or time values changed in every measure; total number of beats increased from 25 to 31.

Used in Experimental Test II-RR as Item No. 11

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.35	76
3rd grade	.05	.49*	47 58
4th grade		.35	60
5th grade		.44*	63
In Test V			
	School		
1st grade	A	.34	76
	B	.55*	48
	A+B		66
2nd grade	A	.23	41
	B	.38*	65
	C	.07	76
	E	.30	56
	A+B+C+E		63
3rd grade	C	.07	70
	E	.64*	62
	C+E	.	67
4th grade	B	.48	81
	D	.39*	72
	B+D		74
5th grade	A	.19	59
	C	.07	63
	E	.26	72
	A+C+E		64
All grades	All	.26*	64

Composer Jan Sibelius — Item No. for Test V - 7

Title Voces Intimae from String Quartet in D minor — Recording Label: 1st version S 10 C; 2nd version S 10 A

Playing time 29 sec. Key d minor Tempo Moderate (Allegretto)

Element Changed Rhythm Medium String Quartet

Nature of Mutilation Melodic rhythm made up entirely of quarter notes and eighth notes is changed to a confused grouping of half, quarter, eighth and sixteenth notes. Accompanying half note chords coming regularly on the first beat of each measure are either lengthened or shortened in duration and three are displaced from their first beat positions.

Used in Experimental Test II-RR as Item No. 13

	School	Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.28	46
3rd grade		.46* .21	70 73
4th grade		.40	54
5th grade		.45*	67
In Test V			
1st grade	A	.35*	64
	B	.41*	52
	A+B		59
2nd grade	A	.57	52
	B	-.16	73
	C	.21	76
	E	.49*	67
	A+B+C+E		65
3rd grade	C	.04	75
	E	.18	58
	C+E		68
4th grade	B	.52	86
	D	.29	86
	B+D		86
5th grade	A	.15	67
	C	.34	83
	E	.24	80
	A+C+E		74
All grades	All	.29*	69

S-10-A

Voces intimae - Quartet in D minor

Sibelius

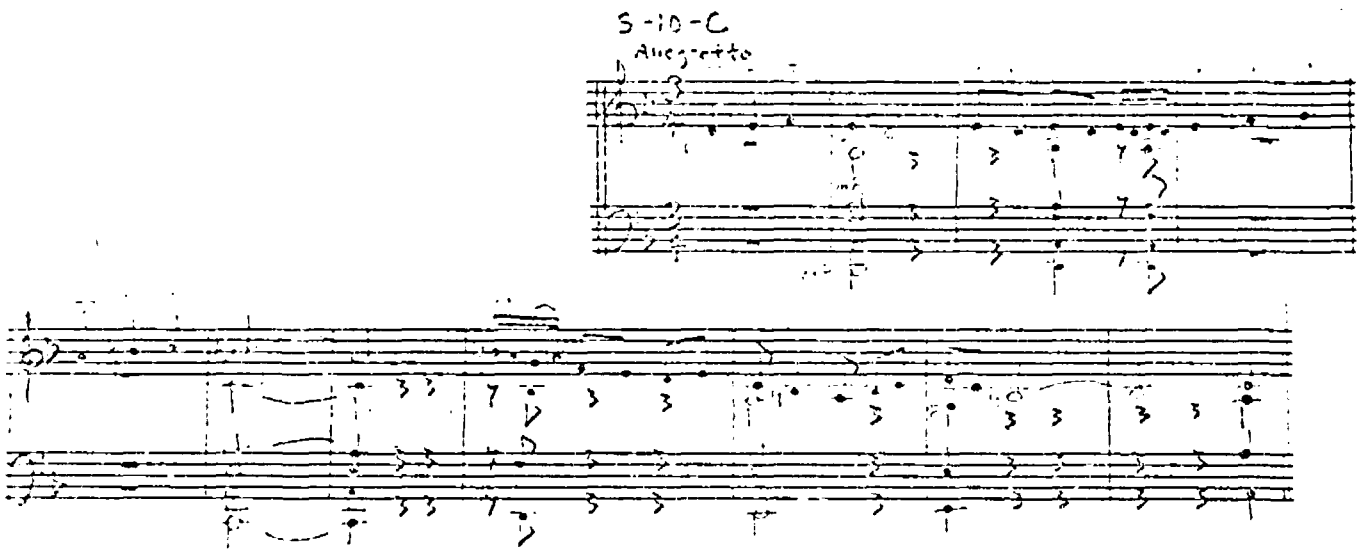


Figure 2. Score and Mutilation for Item 7. An example of the melody and the accompaniment both being given altered time values. Moderate discriminating power.

Composer

Item No. for Test V - 8

Title Jingle Bells

Recording Label
1st version Q 31 D
2nd version Q 31 A

Playing time 13 sec. Key Db

Tempo Fast

Element Changed Harmony

Medium Organ

Nature of Mutilation: Instead of a left hand arpeggiated figure on the tonic chord, the left hand has, in the second measure, a V7 chord with raised 5th and raised 7th leading to a descending chromatic scale for the third and fourth measures. The left hand has a dissonant chord on the last beat.

Used in Experimental Test III-H as Item No. 13

		Discrimination Index		Difficulty Percent responding correctly	
In Experimental Test					
2nd grade			.62*		75
3rd grade		.18	.56	70	90
4th grade			.70		85
5th grade					
In Test V					
	School				
1st grade	A		.33		86
	B		.62		84
	A+B				85
2nd grade	A		.13		79
	B		.43		84
	C		.27		82
	E		.32		78
	A+B+C+E				80
3rd grade	C		.26		93
	E		-.15		79
	C+E				88
4th grade	B		.00		100
	D		.31		92
	B+D				94
5th grade	A		.20		93
	C		.46		83
	E		.79		93
	A+C+E				91
All grades	All		.35		87

Composer Franz Joseph Haydn

Item No. for Test V - 9

Title Partida in F

Recording Label
1st version W 32 A
2nd version W 32 B

Playing time 24 sec. Key F

Tempo Moderate

Element Changed Harmony

Medium Woodwind Quintet

Nature of Mutilation: In the second, fourth and sixth measures the flute is raised one half-step, the clarinet is lowered one half-step, the horn is lowered one whole step and the bassoon is raised one whole step, while the oboe, which is the highest sounding and has the melody, is unchanged.

Used in Experimental Test III-H as Item No. 12

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.59*	65
3rd grade	31.	.53	85 90
4th grade		.66*	60
5th grade			
In Test V			
	School		
1st grade	A	.25	71
	B	.62*	72
	A+B		71
2nd grade	A	.36	69
	B	.39	77
	C	.44*	71
	E	.04	44
	A+B+C+E		64
3rd grade	C	.45*	70
	E	.24	83
	C+E		74
4th grade	B	.21	76
	D	.40	91
	B+D		87
5th grade	A	.14	97
	C	.56	77
	E	.74	85
	A+C+E		89
All grades	All	.43	76

W-32-A Partita in F - 3rd mvt. Haydn

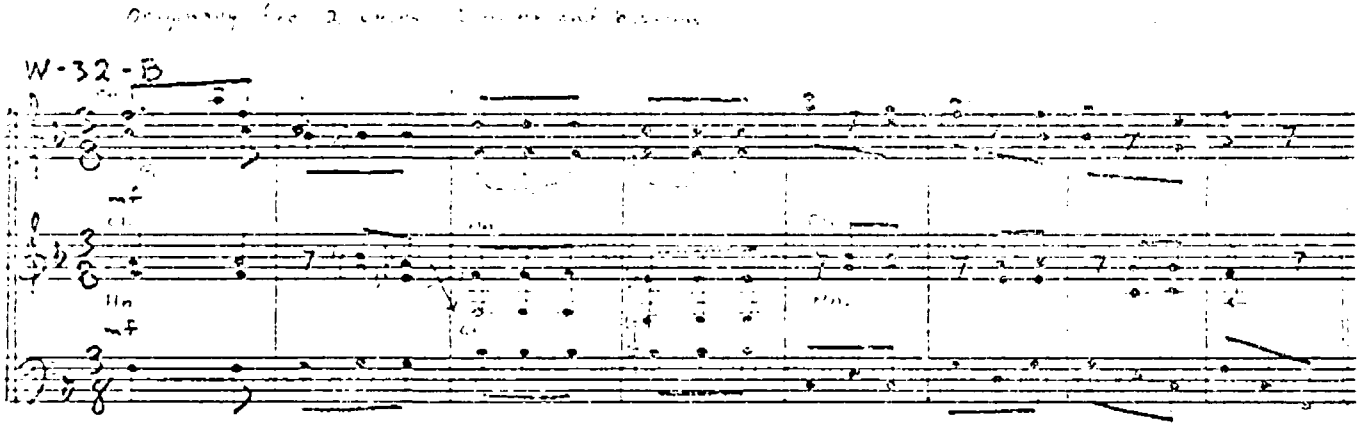


Figure 3. Score and mutilation for Item 9. This is an example of inconsistency of style. The odd numbered measures retain their classical harmony unchanged; even number measures have sharp and illogical dissonances. Item has good discriminating power, but is almost too easy.

Composer — Item No. for Test V - 10

Title Volksleider No. 18 — Recording Label
1st version S 24 B
2nd version S 24 A

Playing time 19 sec. Key C — Tempo Moderate

Element Changed Harmony — Medium String Quartet

Nature of Mutilation: Instead of continuing the C chord through the first measure, the arranger substituted an Ab-Bb-C-D cluster in open position followed in the second measure by a diminished seventh chord in place of a G (V) chord, etc.

Used in Experimental Test III-H as Item No. 14

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.30	65
3rd grade	.12	.43*	75 68
4th grade		.57*	70
5th grade			
In Test V			
	School		
1st grade	A	.26	67
	B	.72	76
	A+B	.	69
2nd grade	A	.22	79
	B	.35	85
	C	.59	82
	E	-.05	81
	A+B+C+E		81
3rd grade	C	.47*	61
	E	.06	71
	C+E		64
4th grade	B	.15	50
	D	.31	78
	B+D		81
5th grade	A	.31	79
	C	.30	70
	E	.01	85
	A+C+E		78
All grades	All	.29*	76

Composer Domenico Scarlatti — Item No. for Test V - 11

Title Toccata Quarto
Minuetto

Recording Label
1st version P 13 E
2nd version P 13 A

Playing time 14 sec. Key e minor Tempo Fast

Element Changed Harmony Medium Piano

Nature of Mutilation: The four-note e minor scale passage followed first by a dominant arpeggio and then by a subdominant arpeggio in the left hand is replaced with three- and two-note chords which abruptly modulate to C, c minor, D, C, F and C and an ending on a b minor triad instead of on V^7 in e minor.

Used in Experimental Test III-H as Item No. 11

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.66*	75
3rd grade		.31	90
4th grade	.48*		65
5th grade		.71*	65
In Test V			
	School		
1st grade	A	.17	48
	B	.01	24
	A+B		40
2nd grade	A	.32	52
	B	-.35	42
	C	.19	29
	E	.36	52
	A+B+C+E		45
3rd grade	C	.28	57
	E	.02	42
	C+E		52
4th grade	B	.19	57
	D	.50	77
	B+D		72
5th grade	A	.29*	71
	C	.09	77
	E	.50*	70
	A+C+E		72
All grades	All	.28*	55

Composer Max Reger — Item No. for Test V - 12

Title Scherzo No. 5 — Recording Label
1st version P 20 A
2nd version P 20 B

Playing time 21 sec. Key C — Tempo Fast

Element Changed Harmony — Medium Piano

Nature of Mutilation: In measures one to four a low discord, C-F-G-B) repeated four times replaces mid-keyboard 2-note chords (I, IV, V, and I). In measures five and six the left hand is a seventh instead of an octave below the right hand.

Used in Experimental Test III-H as Item No. 12

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.13	60
3rd grade		.62* .49*	60 71
4th grade		.65	85
5th grade			
In Test V			
	School		
1st grade	A	.06	42
	B	.17	44
	A+B		43
2nd grade	A	.03	52
	B	.35	31
	C	-.18	41
	E	.24	56
	A+B+C+E		
3rd grade	C	.18	55
	E	.29	46
	C+E		
4th grade	B	.36	62
	D	.38	83
	B+D		
5th grade	A	.41	83
	C	.59	77
	E	.16	77
	A+C+E		
All grades	All	.31*	58

Composer

Item No. for Test V - 13

Title America

Recording Label
1st version Q 33 A
2nd version Q 33 B

Playing time 24 sec. Key F

Tempo Moderate

Element Changed Harmony

Medium Organ

Nature of Mutilation: The harmony is kept fairly consonant, but it is allowed to modulate rapidly three times, but with returns to key of F each time. For example, in the first measure in place of two F chords and a g minor one in first inversion, there is an F chord, a Db triad in second inversion and an e minor triad in open harmony.

Used in Experimental Test III-H as Item No. 4

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.54*	75
3rd grade		.00	100
4th grade		.00	100
5th grade		.51	90
In Test V			
	School		
1st grade	A	.26	88
	B	.47	88
	A+B		88
2nd grade	A	.27	97
	B	.40	92
	C	.57*	71
	E	.17	74
	A+B+C+E		64
3rd grade	C	.40	89
	E	.05	83
	C+E		87
4th grade	B	.00	100
	D	.21	98
	B+D		98
5th grade	A	.23	97
	C	.00	100
	E	.50	97
	A+C+E		98
All grades	All	.34	91

Composer Dominic Zipoli

Item No. for Test V - 14

Title Gavotta

Recording Label
1st version P 4 B
2nd version P 4 A

Playing time 23 sec. Key D

Tempo Moderate

Element Changed Harmony

Medium Piano

Nature of Notilation : "Borrowed chords" are used to depart from the traditional tonality structure.

Used in Experimental Test III-H as Item No. 5

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.55*	55
3rd grade	.23	.25	70 93
4th grade		.18	95
5th grade			
In Test V			
	School		
1st grade	A	.13	88
	B	.52*	68
	A+B		81
2nd grade	A	-.15	90
	B	.28	88
	C	.50	76
	E	.14	76
	A+B+C+E		83
3rd grade	C	.33*	73
	E	.00	79
	C+E	.	75
4th grade	B	.34	95
	D	.52	91
	B+D		92
5th grade	A	.17	90
	C	.45	87
	E	.32	90
	A+C+E		89
All grades	All	.28	84

Composer W. A. Mozart

Item No. for Test V - 15

Title Rondo

Recording Label
1st version P 12 B
2nd version P 12 A

Playing time 25 sec. Key D

Tempo Fast

Element Changed Harmony

Medium Piano

Nature of Mutilation: At the beginning of the second, third and fourth measures the left hand part is transposed up a half step, a whole step, and a minor third respectively. After that Harmony appearing a tenth below the melody is replace by harmony a sixth below the melody which necessitates modulation to remote keys.

Used in Experimental Test III-H as Item No. 6

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.50*	75
3rd grade		.37	81
4th grade	.29	.70	80 85
5th grade			
In Test V			
	School		
1st grade	A	.09	43
	B	.15	64
	A+B		51
2nd grade	A	.34	45
	B	.09	69
	C	-.23	59
	E	.14	56
	A+B+C+E		56
3rd grade	C	.22	70
	E	.19	42
	C+E		60
4th grade	B	.18	95
	D	.54	89
	B+D		90
5th grade	A	.14	84
	C	.54*	70
	E	.07	88
	A+C+E		82
All grades	All	.26*	67

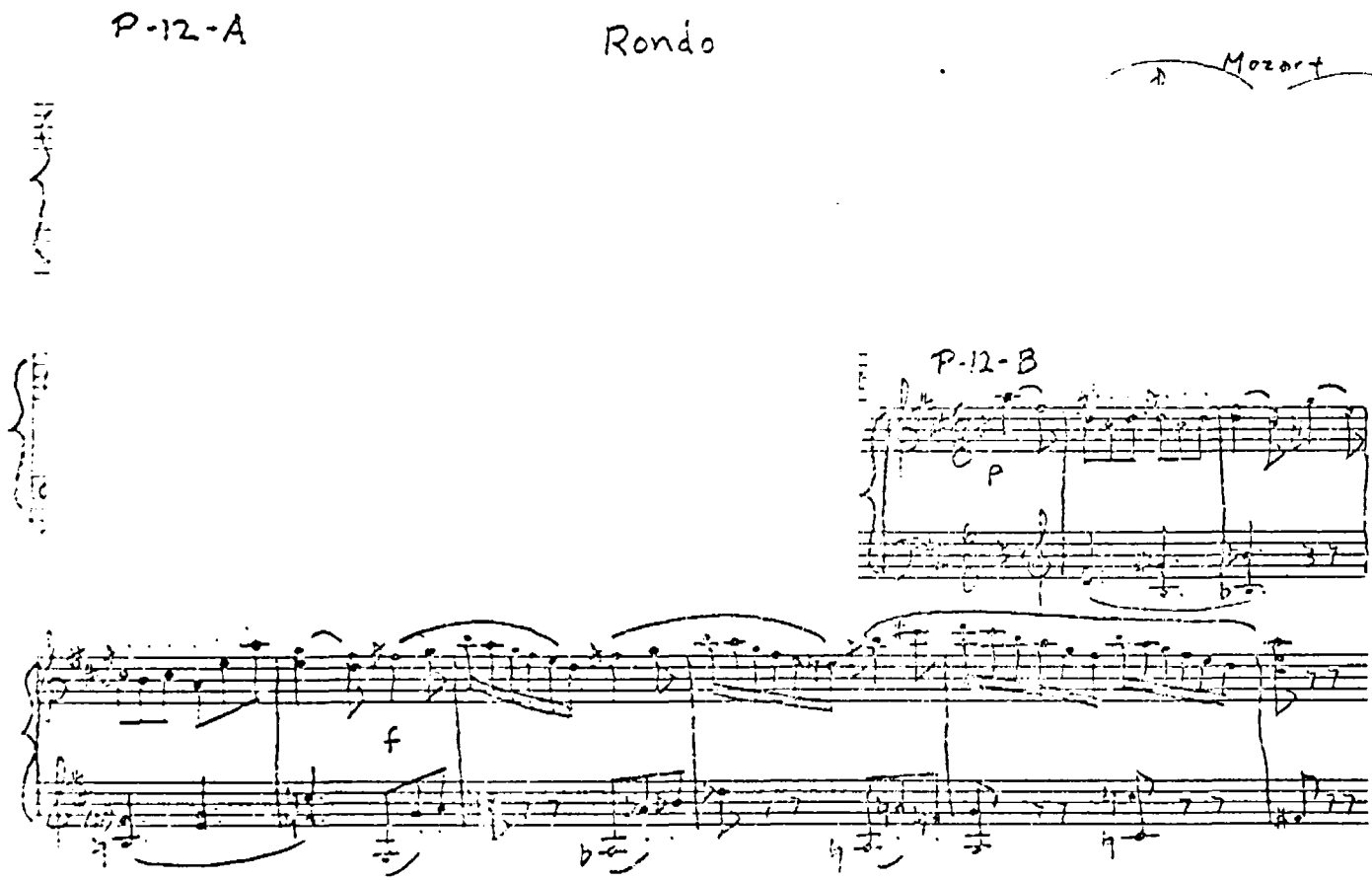


Figure 4. Score and Mutation for Item 15. An example of harmonic change that makes music more dissonant and polytonal. This is one of the weakest of the harmony items in discriminating power, but it was still rated "good" by the computer.

Composer Serge Prokofieff — Item No. for Test V - 16

Title Sonata No. 2, 4th mvt. — Recording Label: 1st version XP 20 A; 2nd version XP 20 B

Playing time 24 sec. Key d minor Tempo Fast

Element Changed Melody Medium Piano

Nature of Mutilation: Whereas the original melody began:
1 5 1 5 8 5 7 4 6 3 the altered version began:
1 5 1 8 8 5 8 5 8 5 The melody for the second half of the item is also made more repetitious.

Used in Experimental Test (1964) as Item No.

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.45*	67
3rd grade	.42	.36	87 65
4th grade		.36	78
5th grade		.31	85
In Test V			
	School		
1st grade	A	.15	74
	B	.21	64
	A+B		70
2nd grade	A	.27	74
	B	.28	81
	C	.12	94
	E	.36	63
	A+B+C+E		76
3rd grade	C	.08	68
	E	.26	54
	C+E		62
4th grade	B	.25	43
	D		
	B+D		
5th grade	A	.13	86
	C	.22	60
	E	.32*	40
	A+C+E		65
All grades	All	.19	68

Composer Volkslieder

Item No. for Test V - 17

Title Volkslieder No. 19

Recording Label
1st version W 15 B
2nd version W 15 A

Playing time 20 sec. Key Ab

Tempo Slow

Element Changed Melody

Medium Woodwind Quintet

Nature of Mutilation: Melodic range is compressed. Whereas the original tune is made entirely from six scale steps, the deformation has only five and in the first six of the eight measures there only three melody pitches in the mutilation to six in the original.

Used in Experimental Test IV-M as Item No. 6

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.10	79
3rd grade		.43* .27	53 56
4th grade		.58*	57
5th grade		.41	52
In Test V			
	School		
1st grade	A	.36	79
	B	.60*	68
	A+B		75
2nd grade	A	.33	79
	B	.30	62
	C	.66*	71
	E	.50	78
	A+B+C+E		72
3rd grade	C	.30	68
	E	.33	67
	C+E		68
4th grade	B	.29	62
	D		
	B+D		
5th grade	A	.23	81
	C	.30	73
	E	.58*	70
	A+C+E		76
All grades	All	.37*	73

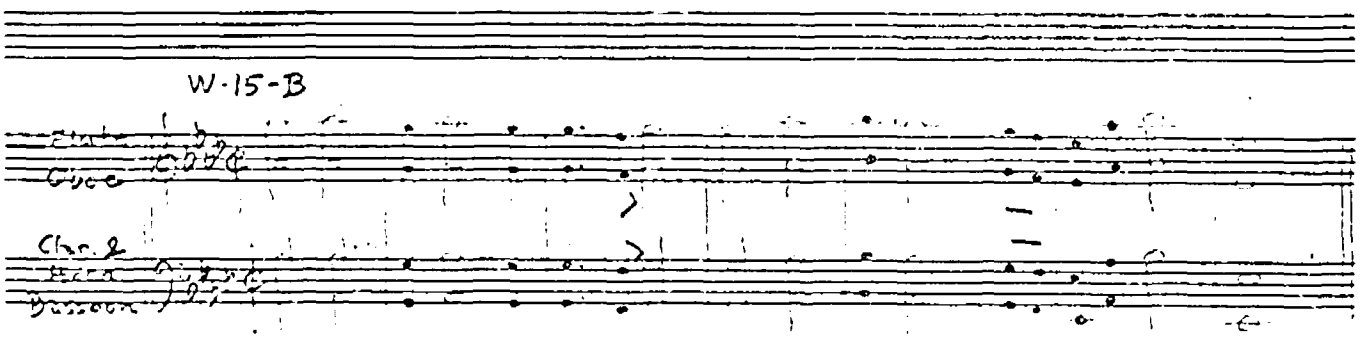


Figure 5. Scores for item 17. This is an example of a melody-only item which, among the melody-changed items, has strong discriminating power. The mutilation has slightly less range and is more monotonous than the original melody.

Composer Ludwig von Beethoven — Item No. for Test V - 18

Title Sonata Op. 27, No. 1, 2nd theme — Recording Label: 1st version XP 44 A; 2nd version XP 44 B

Playing time 25 sec. Key C — Tempo Fast

Element Changed Melody — Medium Piano

Nature of Mutilation : The second half of the phrase deteriorates into a rapid alternation of two tones.

Used in Experimental Test IV-M as Item No. 7

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade		.56*	50
3rd grade		.54* .69*	54
4th grade		.48*	48
5th grade		.20	67
In Test V	School		
1st grade	A	.32*	64
	B	.20	48
	A+B	.21	58
2nd grade	A	.29	62
	B	.44*	46
	C	.83	76
	E	.34	44
	A+B+C+E	.38*	55
3rd grade	C	.43*	70
	E	.38	67
	C+E	.37*	69
4th grade	B	.46*	67
	D		
	B+D		
5th grade	A	.51*	57
	C	.34	43
	E	.52*	30
	A+C+E	.42*	45
All grades	All	.30*	56

Composer G. F. Handel

Item No. for Test V - 19

Title Fugue Op. 3a, No. 3 (subject only)

Recording Label
1st version Q 11 B
2nd version Q 11 A

Playing time 24 sec. Key Bb

Tempo Moderate

Element Changed Melody

Medium Organ

Nature of Mutilation: The mutilated version contains more repeated tones and repeated patterns of two eighth notes.

Used in Experimental Test IV-M as Item No. 8

		Discrimination Index		Difficulty Percent responding correctly
In Experimental Test				
2nd grade		.41		46
3rd grade		.23 .42	33	50
4th grade		.45*		57
5th grade		.31		56
In Test V				
	School			
1st grade	A	.33*		67
	B	.14		56
	A+B	.26*		62
2nd grade	A	.10		62
	B	.15		77
	C	.31		76
	E	-.14		74
	A+B+C+E	.00		71
3rd grade	C	.62*		66
	E	.45*		58
	C+E	.42*		63
4th grade	B	-.05		67
	D			
	B+D			
5th grade	A	.30*		64
	C	.07		57
	E	.46*		60
	A+C+E	.29*		61
All grades	All	.26*		65

Composer G. F. Handel — Item No. for Test V - 20

Title Fugue Op. 3c, No. 9 (subject only) — Recording Label: 1st version Q 10 B; 2nd version Q 10 A

Playing time 22 sec. Key b minor Tempo Moderate

Element Changed Melody Medium Organ

Nature of Mutilation: Downward skip of a fifth reduced to a third; upward skip of a sixth reduced to a third; and downward skip of a fourth followed by two descending seconds replaced by three repeated notes and a drop of a fourth. Subject ends on the fifth scale step instead of on the tonic.

Used in Experimental Test IV-M as Item No. 15

		Discrimination Index		Difficulty Percent responding correctly
In Experimental Test				
2nd grade		.52*		58
3rd grade		.49*	60	50
4th grade		.51*		52
5th grade		.44*		74
	.42*			
In Test V				
	School			
1st grade	A	.24*		43
	B	.13		36
	A+B	.22		40
2nd grade	A	.63*		48
	B	.31		58
	C	.26		24
	E	.10		59
	A+B+C+E	.23		49
3rd grade	C	.28		57
	E	.16*		33
	C+E	.02		48
4th grade	B	.17		52
	D			
	B+D			
5th grade	A	.30*		47
	C	.23		57
	E	.10		50
	A+C+E	.17		50
All grades	All	.11		48

B

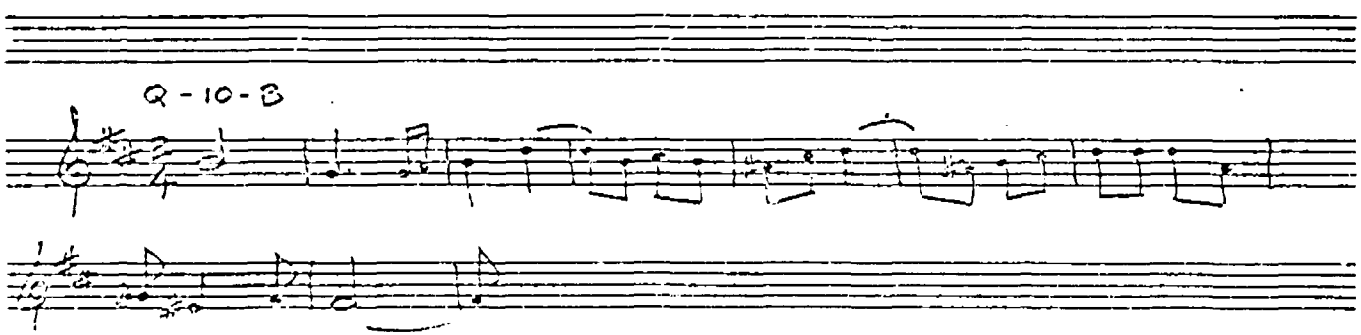


Figure 6. Music for Item 20. This pair had very low discriminating power, the difference in melodic strength being too subtle.

Composer H. Zilcher

Item No. for Test V - 21

Title Sleigh Bells

Recording Label
1st version XP 49 A
2nd version XP 49 B

Playing time 19 sec. Key a minor Tempo Moderate

Element Changed Melody Medium Piano

Nature of Mutilation: The climax in the phrase is eliminated by replacing the highest note with one a fifth lower and replacing the descending scale line with a trill movement in the same time values.

Used in Experimental Test IV-M as Item No. 16

		Discrimination Index		Difficulty Percent responding correctly
In Experimental Test				
2nd grade		.36		62
3rd grade	.49	.64*	47	58
4th grade		.19		43
5th grade		.28		52
In Test V				
	School			
1st grade	A	-.13		33
	B	-.06		36
	A+B	.02		34
2nd grade	A	.23		52
	B	.11		65
	C	.23		29
	E	.09		52
	A+B+C+E	.32*		51
3rd grade	C	.01		36
	E	.15		58
	C+E	.03		44
4th grade	B	.41		48
	D			
	B+D			
5th grade	A	.30*		47
	C	.34		43
	E	.48*		53
	A+C+E	.40*		48
All grades	All	.18		46

Composer Henry Purcell — Item No. for Test V - 22

Title Suite I - Prelude — Recording Label
1st version PX 85 B
2nd version PX 85 A

Playing time 44 sec. Key — Tempo Moderate

Element Changed Melody — Medium Piano

Nature of Mutilation : Melodic 4-note motifs changed from arpeggio-like shapes to "doddles" between two tones.

Used in Experimental Test IV-M as Item No. 17

		Discrimination Index		Difficulty Percent responding correctly
In Experimental Test				
2nd grade		.42*		71
3rd grade		.33 .45*	63	69
4th grade		.40		78
5th grade		.09		74
In Test V				
	School			
1st grade	A	.05		45
	B	.27		40
	A+B	.28*		43
2nd grade	A	.40		76
	B	.33		73
	C	-.28		47
	E	-.13		56
	A+B+C+E	.23		64
3rd grade	C	.20		61
	E	.41		42
	C+E	.36*		54
4th grade	B	.07		62
	D			
	B+D			
5th grade	A	.39*		74
	C	-.09		80
	E	.19		70
	A+C+E	.21		74
All grades	All	.26*		62

Composer G. F. Handel

Item No. for Test V - 23

Title Concerto No. 8
Allemande

Recording Label
1st version S 32 A
2nd version S 32 B

Playing time 42 sec. Key c minor Tempo Slow

Element Changed Harmony Medium String Quartet

Nature of Mutilation; The inner parts, second violin and viola, on alternate two-note patterns are written one note higher in the second violin part and two notes higher in the viola part.

Used in Experimental Test III-H as Item No. 9

		Discrimination Index		Difficulty Percent responding correctly
In Experimental Test				
2nd grade		.59*		65
3rd grade		.31 .53	85	90
4th grade		.66*		60
5th grade				
In Test V				
	School			
1st grade	A	.24		55
	B	.38		60
	A+B	.23		57
2nd grade	A	-.09		55
	B	.21		50
	C	.22		65
	E	.03		48
	A+B+C+E	-.01		54
3rd grade	C	-.14		57
	E	.30		50
	C+E	-.01		54
4th grade	B	.16		57
	D			
	B+D			
5th grade	A	.20		76
	C	.31		60
	E	.34*		57
	A+C+E	.27*		66
All grades	All	.22		59

Composer Johan Pez

Item No. for Test V - 24

Title Bouree

Recording Label
1st version S 38 A
2nd version S 38 B

Playing time 30 sec. Key d minor Tempo Moderate

Element Changed Melody Medium String Quartet

Nature of Mutilation: The melody is reduced in compass from a tenth to an octave and in measures three through six it is reduced to the range of a third from the range of a tenth.

Used in Experimental Test I as Item No. 15

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade			
3rd grade		.23	21
4th grade			
5th grade			
In Test V			
	School		
1st grade	A	-.03	35
	B	.35	44
	A+B	.15	36
2nd grade	A	.29	76
	B	.12	46
	C	-.30	29
	E	.22	44
	A+B+C+E	.22	51
3rd grade	C	.28	43
	E	.02	42
	C+E	.27*	43
4th grade	B	.11	33
	D		
	B+D		
5th grade	A	.21	40
	C	.19	47
	E	.20	48
	A+C+E	.19	44
All grades	All	.16	44

Composer J. C. Bach — Item No. for Test V - 25

Title Quintet No. 3 — Recording Label
1st version W 2 B
2nd version W 2 A

Playing time 24 sec. Key Eb — Tempo Fast

Element Changed Harmony — Medium Woodwind Quintet

Nature of Mutilation: The double tonic and dominant pedal point in the first four measures is changed to chords, I - VII - V - IV. Where the flute melody has been harmonized a third lower by oboe, it is harmonized a sixth lower and visa versa. I - V^7 - I in the next to last measure is changed to parallel chords of a fourth and VI^7 with raised third. The mutilation modulates twice.

Used in Experimental Test I as Item No. 7

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade			
3rd grade		32.*	41
4th grade			
5th grade			
In Test V	School		
1st grade	A	-.02	35
	B	.41	64
	A+B	.22	46
2nd grade	A	.27	35
	B	-.04	53
	C	-.02	24
	E	-.02	44
	A+B+C+E	.21	54
3rd grade	C	.39*	61
	E	.19	42
	C+E	.37*	54
4th grade	B	-.01	48
	D		
	B+D		
5th grade	A	.08	72
	C	.19	87
	E	.01	75
	A+C+E	.09	76
All grades	All	.22	60

Composer Beethoven — Item No. for Test V - 26

Title Quartet No. 7, Op.59, No. 1 — Recording Label
1st version S 4 A
2nd version S 4 B

Playing time 33 sec. Key F — Tempo Moderate

Element Changed Melody — Medium String Quartet

Nature of Mutilation: In the melody, which is in the cello, the first, fourth, fifth, sixth and eighth measures start with the same note in both versions, but the intervening notes are either in different order or are changed entirely - without change of key. Melodic range is a tenth in both versions, but the highest and lowest notes in the mutilation are one note lower than in the original.

Used in Experimental Test 1 as Item No. 8

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade			
3rd grade		.44*	.35
4th grade			
5th grade			
In Test V			
	School		
1st grade	A	.09	48
	B	.30	56
	A+B	.18	51
2nd grade	A	.26	62
	B	.28	38
	C	.10	29
	E	.12	52
	A+B+C+E	.18	47
3rd grade	C	.11	57
	E	-.11	63
	C+E		59
4th grade	B	.01	59
	D		
	B+D		
5th grade	A	.31*	74
	C	.10	60
	E	.21	55
	A+C+E	.18	65
All grades	All	.19	56

Composer G. Merkel Item No. for Test V - 27

Title Andante Op. 122, No. 1 Recording Label
1st version Q 34 A
2nd version Q 34 B

Playing time 43 sec. Key Ab Tempo Slow

Element Changed Rhythm Medium Organ

Nature of Mutilation: Middle voice is changed from beat to off-beat and bass line rhythm is displaced.

Used in Experimental Test I as Item No. 11

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade			
3rd grade		.44*	36
4th grade			
5th grade			
In Test V			
	School		
1st grade	A	.29	38
	B	.26	56
	A+B	.24	45
2nd grade	A	.22	59
	B	.41	65
	C	.27	82
	E	-.04	56
	A+B+C+E	.18	63
3rd grade	C	.14	43
	E	.49*	46
	C+E	.25*	44
4th grade	B	.10	71
	D		
	B+D		
5th grade	A	.19	63
	C	.45*	50
	E	.23	80
	A+C+E	.24	65
All grades	All	.28*	58

Composer Bertold Hummel

Item No. for Test V - 28

Title Bläser Quintet
3rd movement - Burleske

Recording Label
1st version W 23 B
2nd version W 23 A

Playing time 28 sec. Key atonal Tempo Fast

Element Changed Harmony Medium Woodwind Quintet

Nature of Mutilation: Repeated note figures in seconds between oboe and bassoon are changed to sixths, those in 9ths between flute and clarinet are changed to fifths and oboe is added to complete a diminished seventh chord in open harmony.

Used in Experimental Test I as Item No. 13

		Discrimination Index	Difficulty Percent responding correctly
In Experimental Test			
2nd grade 3rd grade 4th grade 5th grade		.28*	27
In Test V			
	School		
1st grade	A	.28	50
	B	.33	44
	A+B	.29*	48
2nd grade	A	.52*	52
	B	.39	62
	C	-.02	58
	E	.36	59
	A+B+C+E	.32*	57
3rd grade	C	.11	55
	E	.34	63
	C+E	.15	
4th grade	B	.42*	67
	D		
	B+D		
5th grade	A	.33*	67
	C	.58	80
	E	.45*	70
	A+C+E	.40*	71
All grades	All	.37*	61

W-23-A

Bläser Quintet Fl. Berfold Hummel

63

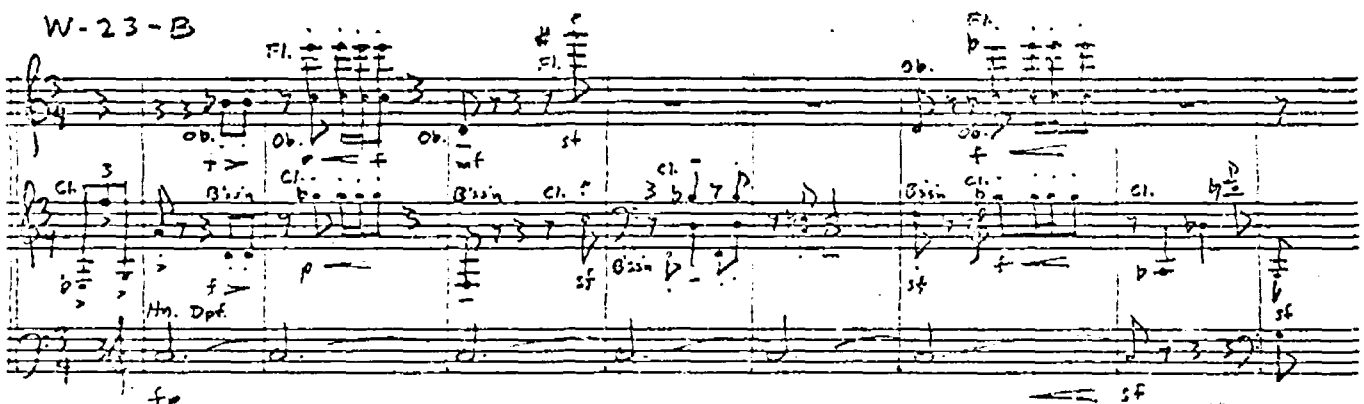


Figure 7. Score and Mutilation for Item 28. An example in which the harmony is made less dissonant in the deformation. Item has fairly high discriminating power.

Thirteen of the 28 items fully met the computer program's criteria for useful, discriminating items. These were items 1, 5, 6, 7, 11, 12, 15, 17, 18, 19, 22, 27 and 28.

Since the test was designed for lower elementary school children, for whom it was assumed there was a need for easy items to motivate pupils to listen intently to the test, it might be desirable to include items which more than 75% of the children marked correctly. It the cut-off point for easy items were moved to 80% giving the correct answer instead of 75%, three more items, 3, 9 and 10 would have high enough discrimination indices to qualify. If the cut-off point were moved to 85%, two additional items, 4 and 14 would meet the criteria.

This leaves ten items of questionable value, 2, 8, 13, 16, 20, 21, 23, 24, 25 and 26 for which replacements should be sought. Of these only one is a rhythm-changed item, four have harmonic changes and five have melodic modifications.

The average difficulty of items classified as slow in tempo was 65.5% (getting items correct); of items moderate in tempo, 65.9%; and of fast pieces, 64.7%. The effect of tempo on item difficulty appeared minimal.

With regard to the medium of performance, piano items proved the most difficult with 62.9% correct responses, whereas the difficulty index for string items was 63.5%, for woodwind items 67.2% and for organ 68.3%.

The average difficulty for rhythm items was 66.3%, for harmony items 69.8% and for melody items 57.9%. It was more difficult then, for children to judge the rightness of the melody than the rightness of either the harmony or the rhythm in these 28 items.

The difficulty of items in which the correct version of a piece was played first was 62.4%, while 67.6% of the responses were correct when the mutilated version was played first.

The most difficult items were 2 and 20. In number 2 the melody was not altered, but only the rhythm of the chords beneath. Number 20 was an unaccompanied organ fugue subject, the mutilated version of which was not greatly inferior in melodic interest to the original. Both of these difficult items could be characterized as having been too subtly altered.

The easiest items were 13 and 8, "America" and "Jingle Bells," the only familiar tunes in the test and both harmonically distorted. Percent of pupils making correct responses on these two items were 87 and 91 respectively. While more discriminating items might replace these two, still it might be advantageous to retain them for their psychological effect in giving pupils confidence.

The correlation between the length of an item and its difficulty, computed by the product-moment method, was .26, not significant at the 10% level of confidence.

The breakdown of the 18 acceptable items as to stylistic period and performing medium is shown in Figure 6.

TABLE 6. NUMBER OF USABLE ITEMS IN FORM V BY PERIOD AND PERFORMING MEDIUM.

Medium	Baroque	Classical	Romantic	Modern or Contemporary
Piano	3	2	2	1
Organ	1		2	
String Quartet	2		1	
Woodwind Quintet	1	1		2

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Item analysis had been used to identify the best items in the Experimental Tests I, II-RR, III-H, and IV-M. It was anticipated that when these "best" items were assembled into Test V that, in combination, they would continue to perform well and that the reliability and discriminating power of Form V would emerge at high levels.

The performance of Test V when given 383 pupils in grades 1 through 5 in various schools was disappointing in that reliability coefficient obtained for Test V were not higher or more consistent than those obtained in the Experimental Tests and some items did not function as effectively as anticipated.

While very satisfactory reliability was obtained for one first grade class, the mean score for first grade was only 16.3, which is not very much above a score which might be obtained by guessing on a 28-item 2-choice test. If we ignore the two familiar too-easy items, "America" and "Jingle Bells," we have a mean score of 14.3 on a 26-item test. This is uncomfortably low and would tend to discredit the test until we take a look at the scores themselves. Of the 67 first graders, one had 25 correct responses, two had 22, one had 21, and five had 20 correct responses. It is obvious that these pupils, although they are only 13% of the group, are able to make judgments about music and Test V is measuring this ability.

The variation in the discrimination power of the items for different classes tested was a source of concern. With the same item having high discrimination indices for some classes and low ones (even negative ones) for others, the Test appeared to be functioning too capriciously. The small number of subjects in some of the classes and the differences academically and socio-economically between them may account for some of the wild variation, but additional data on these pupils would be needed to assess this properly. On the other hand, this variation between classes shown in the item analysis may allay some fears about the amount of guessing being done by the respondees. Had guessing been rampant its effect would have been an equalizing one, thus reducing rather than causing the variations observed.

Experimental forms II-RR, III-H and IV-M each required subjects to respond to items with one, and only one, element altered. Judging the "rightness" of items in which melodic changes differentiate the two versions may require a different ability than to make judgments about rhythmically altered items. Consequently

the similarity of responses made on any one of these Experimental Tests may account for the high discrimination indices obtained. When the variously altered items were combined into one test the internal validity of a rhythmically altered item might suffer because the computation compares these responses to a rhythm item with responses on harmony and melody items. This may account for the unsatisfactory performance in Test V of items which looked promising as parts of Experimental Tests II-RR, III-H and IV-M.

We can make only modest claims about the efficacy of the test and the dependability of measurement it affords, but it is our opinion that the general design of Test V is sound and that it can be used to measure in part, at least, the ability of elementary school children to make wholistic aesthetic judgments about concert type music.

At the same time we hasten to admit the imperfections of Test V and the need to further improve it. (See recommendations.)

Errors in Procedure

Two decisions made early in the project may have had adverse effects upon the final results. First, when the distorted versions of the pool of items were being composed (or "decomposed") the possibility that subjects would have to identify which element had been changed had not been eliminated. Consequently, great care was taken to be sure that only one element--rhythm, melody or harmony--was changed in any one mutilation. It is conceivable that items of greater testing power might have been created had this limitation not been observed.

Second, the heavy emphasis on selecting items which had been pre-judged "easy" resulted in an abundance of items with reasonably good discriminating indices that were still too easy to meet the lower limit of test difficulty. Not enough items pre-judged medium difficult were tried out in the Experimental Tests to provide an adequate selection of high discriminating items when it came time to select items for Test V.

Recommendations

Study of the data concerning Test V and of the conclusions presented above leads to the following recommendations:

1. Many more of the items in the recorded pool should be assembled into experimental tests and, after administration to elementary school children, be given item analysis.

2. Ten items, when proven good by the above analysis, should be used to replace in form V the ten items found inferior in discriminating power. Restoration of a balance with regard to stylistic periods and performing media should be a factor in the selection of the ten items.

3. A second test parallel and equivalent to Test V should be developed to facilitate pre-test--post-test uses in research.

4. A testing program should be initiated to standardize the tests resulting from recommendations 2 and 3. A proportionate number of inner-city pupils should be among the pupils tested.

5. A correlation of test scores with scores on the "phrasing" portion of the Gordon Music Aptitude Profile should be undertaken for a measure of external validity. Harrington[24] selected three subtests - melody, tempo and phrasing - from the seven in Gordon Profile and tried them out as a "primary" test with second and third grade pupils. While he concluded that only the rhythm (tempo) and melody subtests should be employed as an aptitude test in grades two and three, he did find that the expression (phrasing) subtest had a reliability of .73 for grade two and .68 for grade three. (Harrington, commenting on this reversal of the common trend for reliability to increase with grade level, a reversal similar to the one found between grades one and two in the present study, had no ready explanation other than that the reversal might be due to "random sampling error.") Harrington's reason for not using the "phrasing" subtest in a primary grades aptitude test was the low correlation the subtest's scores had with teachers' estimates of pupils' skill in expression and phrasing. In spite of this shortcoming this subtest, with simplified directions and increased time for pupil response which Harrington found helpful, would seem better suited at this primary grade level for an external validity correlation than any of the "appreciation" sections of the Wing Measures of Musical Intelligence, the other standardized test with items of a music discrimination nature.

6. Trial runs of the test developed in this project using a three-choice format in which an "in doubt" or "not sure" response can be marked when a pupil does not feel prepared to decide

---

24 Harrington, Charles J., "An Investigation of the Primary Level Musical Aptitude Profile for Use with Second and Third Grade Students," Journal of Research in Music Education, Vol. VII, No. 4, Winter 1969. pp. 359-366.

whether the first version is better than the second or visa versa should be conducted. Such a format might not prove as confusing to first, second and third graders as the one abandoned after being tried in Experimental Test I and Experimental Test II-R wherein identical versions of phrases were to be recognized and marked "same."

7. When Test V has been improved by replacing its poor items with stronger ones, item analysis might be done separately on rhythm items, on harmony items, and on melody items, as well as on the test as a whole.

8. Since there is evidence that different kinds of musical judgments are employed in responding to items in which different elements have been altered, consideration should be given to reporting scores on rhythm items separately, on harmony items separately, and on melody items separately in addition to reporting a composite score.

9. Extension of the present study should be undertaken using jazz, oriental music and electronic music for item material. Experimentation in measuring the ability of young children to judge which of two settings of a song text is the more artistic is another possibility.

10. Based on information theory an analysis of the relation between the effectiveness of items in Test V and the amount of "musical information" contained in each item and its deformation might prove revealing.

11. Other designs for measurement of music discrimination in small children should be sought and tried experimentally.

BIBLIOGRAPHY

Andrews, Frances M., and Deihl, Ned. C., "Development of a Technique for Identifying Elementary School Children's Musical Concepts," Council for Research in Music Education Bulletin, No. 13, Spring 1968, pp. 1-7.

Bridges, Virginia Ann, "An Exploratory Study of the Harmonic Discrimination Ability of Children in Kindergarten through Grade Three in Two Selected Schools," (Ph. D. Dissertation, The Ohio State University, 1965, 65-13,206), reviewed by Kaplan, Barbara Connally, Council for Research in Music Education Bulletin, No. 19, Winter 1970, pp. 30-33.

Carr, Constance, editor, Music for Children's Living, Association for Childhood Education International, Bulletin 96, Washington, D. C., 1955.

Copland, Aaron, What to Listen for in Music, McGraw Hill, 1939.

Crickmore, Leon, "An Approach to Measurement in Music Appreciation," Journal of Research in Music Education, Vol. XVI, No. 3, Fall 1968, pp. 239-253.

Edwards, Allen L., Statistical Methods for the Behavioral Sciences, Rinehart & Company, Inc., New York, 1954.

Greenberg, Marvin, in review of Peterson, Agda Viola, "A Study of Developmental Listening Factors in Children's Ability to Understand Melody," (Dissertation, University of Rochester Eastman School of Music, 1965), in Council for Research in Music Education Bulletin, No. 16, 1968 (?).

Haack, Paul, "A Study of the Development of Music Listening Skills of Secondary School Students," (Dissertation, University of Wisconsin, 1966).

Harrington, Charles J., "An Investigation of the Primary Level Musical Aptitude Profile for Use with Second and Third Grade Students," Journal of Research in Music Education, Vol. VII, No. 4, Winter 1969. pp. 359-368.

Hevner, Kate, "Appreciation of Music and Tests for Appreciation of Music," Studies in College Teaching, Vol. 4, No. 6, pp. 83-151.

Lehman, Paul, Tests and Measurements in Music, Prentice-Hall, Menglewood Cliffs, N. J., 1968.

Lifton, Walter M., "Development of a Music Reaction Test to Measure Affective and Aesthetic Sensitivity," Journal of Research in Music Education, Vol. IX, No. 2, Fall 1961, pp. 157-163.

Long, Newell H., "A Revision of the University of Oregon Music Discrimination Test," (Ed. D. Dissertation, Indiana University, 1965) 65-14,054.

Long, Newell H., "Establishment of Standards for the Indiana-Oregon Music Discrimination Test Based on a Cross-Section of Elementary and Secondary Students with Analysis of Elements of Environment, Intelligence and Musical Experience and Training in Relation to Music Discrimination," U. S. Office of Education Grant No. OEG 3-7-700027-2893, Indiana University, 1969.

Mursell, James L., "Applications of Psychology to the Arts," Teachers' College Record, 37:290-299, Jan. 1936.

Petzold, Robert G., "Auditory Perception by Children," Journal of Research in Music Education, Vol. XVII, No. 1, Spring 1969, p. 85.

Petzold, Robert G., "Development of Auditory Perception of Musical Sounds by Children in the First Six Grades," Journal of Research in Music Education, Vol XI, No. 1, Spring 1963, pp. 21-43.

Pflederer, Marilyn and Sechrest, Lee, How Children Conceptually Organize Music Sounds, Cooperative Research Project No. 5-0256, Northwestern University, 1968, 232 pages.

Pike, Alfred, "The Phenomenological Analysis and Description Experience," Journal of Research in Music Education, Vol. XV, No. 4, Winter 1967, pp. 316-319.

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

Simon, Carol R. and Wohlwill, Joachim F., "An Experimental Study of the Role of Expectation and Variation in Music," Journal of Research in Music Education, Vol. XVI, No. 3, Fall 1968, pp. 227-238.

Taylor, Sam, "Development of Children Aged Seven to Eleven," Journal of Research in Music Education, Vol. XVII, No. 1, Spring 1969, pp. 100-107.

Wing, Herbert D., Tests of Musical Ability and Appreciation, monograph supplement, The British Journal of Psychology, Cambridge University Press, 1948.

Whybrew, William E., Measurement and Evaluation in Music, Wm. C. Brown Company, Dubuque, Iowa, 1962.

APPENDIX A

USO E Music Discrimination Test
Directions for giving form V - RH

(When pupils have been put at ease and each is supplied with a pencil, distribute the answer sheets. Have pupils write or print their names on the blank at the top of the page. Place on the chalkboard:)

1. 1 2
2. 1 2
3. 1 2

Spoken Instructions: If you had two apples and took a bite out of one and then a bite out of the other, you could tell which one tasted better, couldn't you? . . . We are going to taste, by listening to some short pieces of music. You will hear each little piece two times, but one time it won't sound as good as the other time. You are to listen to the piece both times so that you can tell which makes the better music, the first time or the second time. If you think the first time sounds better than the second time, make a circle around the "one" (Pointing to "1" column on a response sheet.) If you think the second time sounds like better music than the first, how do you think it should be marked? (Let a pupil answer, but be sure the class understands "Make a circle around the "2".)

I will help you with the first three pieces. Let's listen to number one. (Play no. 1, versions 1 and 2.)

How many thought the first time sounded better? Raise your hands. How many thought the second time was better? The first time was the better one, so we make a circle around the "1". (Do so on chalkboard.)

Now let us listen to number two. (Play no. 2, both versions.) How many thought the first time sounded better? Hold up your hands. How many thought the second time made better music? Hands? The second time was better, so you should have a circle around the "2". (Do so on chalkboard.)

Here is piece no. 3. (Play both versions.) How many prefer the first time? That's correct. So we should have a circle around which number? (Let class respond.) Around the "1". (Place circle around "1".)

(Examples on chalkboard should look like this:)

1. (1) 2
2. 1 (2)
3. (1) 2

In all three pieces that we have heard, it was the rhythm that was changed.

Now we will hear some more in which the rhythm is not the same. Without any help from me or from anyone else, you are to listen so you can tell which of the two times the music is better. Circle "1" if it sounds better the first time; circle "2" if it sounds better the second time. Get ready. Here is number 4 _____. No. 5_____etc.

(after number 10)

Turn your papers over while we rest our ears.

(Place on chalkboard:)

11. 1 2
12. 1 2

Next we shall hear some pieces in which the harmony has been changed. I will help you with eleven and twelve. Let's listen to number eleven. (Play no. 11.)

The second time sounded better than the first, so we make a circle around which number? - - - - Around the "2". Here is number 12. (Play 12.) Which was the better music, the first time or the second? Which sounded better to you? _____ No. 1 That's correct, so we make a circle around the "1".

(Chalkboard should be marked:)

11. 1 (2)
12. (1) 2

Here are some more pieces in which the harmony has been changed. Pick out which is better, one or two. Number 13_____
No. 14_____etc.

APPENDIX B

SAMPLE ANSWER FORMS

Name _____

TEST V

Part I - Rhythm changed

1. 1 2
2. 1 2
3. 1 2
4. 1 2
5. 1 2
6. 1 2
7. 1 2
8. 1 2
9. 1 2
10. 1 2

Part II - Harmony changed

11. 1 2
12. 1 2
13. 1 2
14. 1 2
15. 1 2
16. 1 2
17. 1 2
18. 1 2
19. 1 2
20. 1 2

Name _____

TEST V

Part III - Melody changed

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

Part IV - Rhythm, harmony or melody changed

10. 1 2
11. 1 2
12. 1 2
13. 1 2
14. 1 2
15. 1 2
16. 1 2
17. 1 2
18. 1 2

APPENDIX C

SCHOOLS AND PERSONNEL ASSISTING PROJECT

Eastern School, Greene County, Ind.	Mrs. Karen Glickman
Edgewood School, Perry Twp., Indianapolis, Ind.	Mrs. Sue Ann Meyer
Ellettsville, Ind., Elementary School	Mrs. Patricia Powell Mrs. Paul Schunn
Monroe County Community Schools Bloomington, Ind.	Dr. Richard Borders Mrs. Joel Mathias Mrs. Bernard Norberg Mrs. Howard Ottensmeyer
Poston Road School, Martinsville, Ind.	Dr. Larry King
Skiles Test School, Lawrence Twp., Indianapolis, Indiana	Mrs. Margaret Weesner
Spencer, Ind., Elementary School	Mrs. Lillian Dyar
University School, Bloomington, Ind.	Mrs. Marcia Hoadley Mrs. Margaret Pounds
Thornton School and Rea School, Vigo County Schools, Terre Haute, Ind.	Mr. Arthur Hill, Jr. Miss Katherine Becker Mr. Leslie Bain
Wayne Twp. School, Bartholomew County, Columbus, Ind.	Mr. Burdell Sell Mrs. Marjorie Fritz