24661315

LIBRARY
Michigan State
University

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 00572 8401

This is to certify that the

dissertation entitled

THE DEVELOPMENT AND VALIDATION OF A COMPUTERIZED DIAGNOSTIC TEST FOR THE PREDICTION OF SUCCESS IN THE FIRST-YEAR MUSIC THEORY SEQUENCE BY INCOMING FRESHMEN AT MICHIGAN STATE UNIVERSITY

presented by

James Peter Colman

has been accepted towards fulfillment
of the requirements for

PhD degree in Music Education

Albert LeBlanc
Major professor
Albert LeBlanc

Date February 15, 1990

MSU is an Affirmative Action/Equal Opportunity Institution O-12771

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
~~NOV 0 5 1992~~ 9730662		

MSU Is An Affirmative Action/Equal Opportunity Institution

THE DEVELOPMENT AND VALIDATION OF A COMPUTERIZED DIAGNOSTIC TEST FOR THE PREDICTION OF SUCCESS IN THE FIRST-YEAR MUSIC THEORY SEQUENCE BY INCOMING FRESHMEN AT MICHIGAN STATE UNIVERSITY

By

James Peter Colman

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

School of Music

1990

ABSTRACT

THE DEVELOPMENT AND VALIDATION OF A COMPUTERIZED DIAGNOSTIC TEST FOR THE PREDICTION OF SUCCESS IN THE FIRST-YEAR MUSIC THEORY SEQUENCE BY INCOMING FRESHMEN AT MICHIGAN STATE UNIVERSITY

By

James P. Colman

Freshmen music students enrolling at many colleges and universities in the United States frequently face a one year music theory course requirement. First-year music theory courses seek to provide all freshmen students with a common theory foundation for the rest of their music training. Some assumptions must be made concerning the present knowledge of incoming students. These assumptions are seldom accurate for all students. The goal of this study was to create a computerized diagnostic test capable of measuring the current music theory achievement of incoming students so that statistical data could replace the assumptions made by college theory professors. Secondly, this study sought to determine whether the newly created test could function as a predictive variable in the evaluation of future success in music theory at Michigan State University.

The test included 90 questions designed from objectives covering all content areas of the first term of the music theory sequence at Michigan State University. The test was implemented on the Macintosh computer using the HyperCard software from Apple Computer, Inc. Each test item included from two to four multiple choice answers. The subjects selected an answer by

clicking with the computer's input device (mouse) on the chosen answer. The computer handled all aspects of the test including administration, data storage, test result printouts, and statistical analysis.

The test was administered to 59 freshmen subjects at the beginning of the Fall term in 1988. The results of the test were correlated with three grade criteria over a period of three college terms: theory lab grades (0-100%), final percentage grades (0-100%), and grade points (0.0-4.0). The test was also correlated with a three-term average computed for all subjects who completed the entire first-year theory sequence. The strongest correlation was found between the test and final grade points. This was surprising since the grade point scale was the least sensitive of the three criteria. The study concluded that the first iteration of the music theory test was sufficiently successful to warrant further development for future use as a diagnostic/predictive tool.

Copyright by

James Peter Colman

1990

Dedicated to my wife who has been

my constant support and encouragement

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to Dr. Albert LeBlanc for his input into my professional development, his willingness to give me opportunities to use my developing research skills, and the tremendous guidance he has provided throughout this endeavor. I am proud to call him my mentor. I also appreciate the substantial input and support of Dr. Charles Ruggiero, Dr. Theodore Johnson, Dr. Corliss Arnold, and the freshman theory students who participated in this study.

Apple®, ImageWriter®, Macintosh®, and HyperCard® are registered trademarks of Apple Computer, Inc.

SuperPaint® is a registered trademark of Silicon Beach Software.

Professional Composer® is a registered trademark of Mark of the Unicorn.

MacRecorder® and SoundEdit® are registered trademarks of Farillon Computing, Inc.

IBM® is a registered trademark of International Business Machines Corporation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - THE PROBLEM

Introduction

Freshmen music students enrolling at many colleges and universities in the United States frequently face a one year music theory course requirement. The course usually consists of two semesters or three quarters. Typically, students enrolling as college music majors have relatively little specific training in music theory. Any theory knowledge they have accumulated they gained through private music lessons or performance in high school band or choir organizations. Financial cut-backs have severely limited the number of high school music theory classes. First-year music theory courses seek to provide all freshmen students with a common theory foundation for the rest of their music training. It is impossible to begin from the very first concepts of music theory training, however; some assumptions must be made concerning the present knowledge of incoming students. These assumptions are seldom accurate for all students. Increased risk of failure results when students attempt to accomplish the requirements of a first-year theory course without the assumed knowledge since they start at a disadvantage. The goal of this study was to create a diagnostic test capable of measuring the current music theory achievement of incoming students so that statistical data could replace the assumptions made by college theory professors. Secondly, this study sought to determine whether the newly created test could function as a predictive variable in the evaluation of future success in music theory at Michigan State University. The end objective of the study was to provide a musically specific test that would give advisors of college music majors another aid for proper advising decisions. Some background in the types of advising problems and educational treatments addressed by this study is in order.

Accurate diagnosis of student abilities and deficiencies is extremely important to successful college and university counseling. Students who receive inadequate advising risk the possibility of incomplete preparation for their chosen profession or even misdirection into a field for which they are ill-suited. Current enrollment policies of United States institutions of higher education permit completely open enrollment, that is, there are no admission requirements, other than available space, hindering a student's acceptance. The policy of open enrollment, however, brings with it the problem of providing each student the most useful education possible while dealing with a myriad of differences in each student's background and needs. Willingham (1974) identified two recent trends which increase the academic advising demands placed upon higher education.

First, a greater diversity of educational alternatives and incentives, including community colleges, federal student aid, and flexible academic programs, has encouraged an influx of new students, particularly minority students, adults, and students previously discouraged from continuing education because of academic weaknesses, resulting in a need for advising flexibility.

Second, economic considerations of students place pressure upon institutions to provide the specific academic need of each student. The stabilization and, in some institutions, decline in student enrollment has greatly increased competition for students. Institutional programs must

accommodate the financial needs of the individual student or the student will look elsewhere. Willingham suggests four classes of treatment for satisfaction of student academic requirements and interests in answer to the demands raised by these trends.

The first treatment places or assigns students to various classes based upon similar abilities or personal characteristics such as similar test scores. The educational techniques may vary among the classes but the subject matter and end objectives should be the same.

A second treatment places students into an instructional sequence on the basis of their current knowledge of the subject. As with the first treatment (assignment), the knowledge of the subject matter and the end objectives are the same for each student, but the student does not invest time in material previously mastered. For example, it is possible that an incoming student might already possess the skills usually developed in the first term of a music theory course. If an accurate assessment of the student's current knowledge were possible, the student could be placed into a subsequent term of music theory.

The third treatment possibility suggested by Willingham, selection, groups students with different ability levels into various instructional programs with different educational content and end objectives. This method is most frequently observed in the offering of advanced classes designed to exceed the usual course content demands and to motivate the student to progress past the normal end objectives for that particular class. An opposite result is possible when students are selected for placement in remedial classes. Students required to take remedial classes might only receive a portion of the material included in the standard class.

Exemption, the last of Willingham's four treatments, excuses students who demonstrate substantial proficiency in a given subject area completing course requirements that emphasize the area of proficiency. Different academic programs apply exemptions differently and several workable strategies are available. The student may or may not receive credit for the exempted classwork, or may have to take another course in place of the exempted class or classes.

Each of the previously discussed treatments has many valid applications. Some applications might necessitate the implementation of more than one treatment. Willingham suggested five methods of testing student abilities to determine proper treatment; proficiency testing, diagnostic testing, evaluation of personal characteristics, aptitude testing, and evaluation of grades. All five methods of evaluation described below were reviewed for this study and diagnostic testing was selected as the most appropriate for achieving the stated goals.

Proficiency testing measures competency in a given course or group of courses. This type of measure may assess factual knowledge, problem solving abilities, or ability to make practical applications as an indication of the extent of the student's knowledge in the tested subject area. The test must only include material taught in the course or courses falling within the scope of the test. Proficiency tests are most useful with placement or exemption treatments.

Diagnostic testing is, in many respects, similar to proficiency testing, but the former provides a more detailed evaluation of what the student knows and what the student does not know. Diagnostic tests are most beneficial when they provide part-scores which allow the test administrator to make accurate assessments of current accomplishments which in turn provide the required information for proper placement of the student.

Evaluation of personal characteristics is a helpful tool when used with selection or assignment treatments although it has no usefulness in the present study. Personal characteristics can include almost any trait not connected to abilities or achievements such as background characteristics, interests, cognitive styles, and attitudes. For example, students might be placed in a participatory class because they have demonstrated greater material retention when allowed to physically interact with items related to the lesson. Evaluation of personal characteristics and interests is perhaps the least useful testing method because it is difficult to produce adequate testing tools. This method is also open to criticism in the area of objective decisions concerning student placement.

Aptitude testing is also helpful as an assessment measure. Aptitudes usually include any cognitive abilities not readily improved through short-term learning. Selection treatment decisions are enhanced by assessment of aptitudes related to general scholastic performance while assignment treatment decisions are enhanced by the assessment of specialized aptitudes.

Finally, a student's high school record offers another assessment tool. Generally, a student's grades provide information on academic performance across a wide range of subject areas including the chosen undergraduate field. The high school record is difficult to interpret, however, because of the lack of standardization. Variance in grading scales, teacher standards, and even the level of academic competition can greatly influence grades and make an evaluation of true accomplishment very difficult.

The previous discussion has examined the need for flexibility in academic guidance because of economic considerations and the influx of new

stu
sug
ass
of t
inf
<u>Ne</u>

inf
fre
acc
(SA
ad
us
st
SA
ex
st
ev
th

sc
ir
m
a
n
a
g

students; the use of assignment, placement, selection, and exemption as suggested treatments which allow for greater flexibility; and the methods of assessment useful in gathering information necessary for correct assignment of treatment types. Attention now turns to the application of this information to the present research study.

Need for the Study

The lack of musically specific and/or standardized academic information available for freshmen students makes the task of advising freshmen music students difficult. On the one hand, college advisors have access to college entrance examinations such as the Scholastic Aptitude Test (SAT) or American College Test (ACT). These standardized tests allow the advisor to make generalized inferences about the advisee's abilities but usually have limited application to specific evaluation of the enrolling student's achievements or aptitudes in his or her major area. Neither the SAT nor the ACT contain sections devoted to musical concepts. The exclusion of such material is not a defect but only a limitation since these standardized tests are designed to provide general information not to evaluate most specific content areas. The general information gathered by these tests is an inadequate basis for advising in music.

College and university advisors may use a student's grades along with scores from standardized tests to make advising choices, but, as the introductory discussion noted, a student's grades are an ambiguous measurement tool at best. Standardized test scores and grades must have additional support from musically specific indicators. Academic advisors need the information produced as the result of development and administration of diagnostic tools specifically designed to enhance more generalized academic indicators. The lack of standardized measurement tools in th

imp

Mic

adv

trac

the

dis

var

wh

me

ins

Mi

the

sec

tes

pr

sc

ap

ev

m

m

te

ir

Ir

p

in the area of music placement and advising at the college level was the main impetus for this study.

The development of a measurement tool specifically for use at Michigan State University could expand the knowledge available to academic advisors at that institution. Designing a test for a specific locale is not without trade-offs. Willingham suggests that "a principal advantage of the local test is the fact that it can be designed for the purpose in mind; the main disadvantage is the fact that the technical quality of locally constructed tests varies a great deal" (p. 27). This study began the process of test development which could eventually culminate in the completion of a useable measurement tool for academic advisors of freshmen music majors at all institutions with content similar to that covered in the theory sequence at Michigan State University. More information on the other applications of the study results is included in the limitations section and the conclusion section. One might raise the objection that there are a number of diagnostic tests for music theory already available. This study stands in contrast to previous test development studies because it is entirely administered and scored by computer.

Numerous articles have been written which describe computer applications in the roles of teacher, drill instructor, test administrator, and evaluator. The expanding number of computer uses in educational areas may provide the solution to one of the tedious aspects of diagnostic measurement, test administration. Prior to the widespread use of computers, test administration required an added expenditure of time and energy by the instructor or added monetary expense to hire someone to supervise the test. Inadequate testing resulted as instructors or departments became unable to provide the necessary time or money.

feas
adm
the
test
gat
Pro

des
inc
wh
the
ou

en
id
ob
be
co
p
ta
F
c
L

l

A secondary goal of this study was to demonstrate the desirability and feasibility of incorporating the available technology in computers into the administration of a diagnostic test and the subsequent statistical analysis of the test results. The use of computers in the administration of a diagnostic test frees the advisor of extra time expenditures by automating information gathering.

Problem Statement

The problem of this study was to complete the initial stages of designing a music theory test capable of diagnosing the current knowledge of incoming college freshmen at Michigan State University and of predicting whether these students would successfully complete the required freshman theory courses. Several subproblems were addressed as the study was carried out.

First, the expectations theory professors placed upon freshmen students entering music study at Michigan State University were defined. Once identified, the chosen content areas were attached to specific behavioral objectives that reflected the expectations of the theory professors. Next, the behavioral objectives were illustrated with test items designed to elicit a correct behavioral response. The test items were then administered to a population sample with the goal of generalizing the results to other samples taken from the same population, incoming freshmen music students. Finally, the results were validated through a test-retest design and correlations were performed to check for reliability.

Definition of Terms

Achievement test - refers to a test designed to evaluate the current levels of knowledge and understanding of a particular content area.

Achiev
readin

of an
musi

when
reco
soun

deve
prog
four
ansv

319
Univ
Mac
vari

ach
the
con

a c
Fa
us

Achievement tests are frequently used at all age levels as an indicator of readiness for promotion.

Aptitude test - refers to a test designed to evaluate the potential abilities of an individual to perform certain skills. Elementary school evaluation of musical aptitudes for playing band instruments is an excellent example.

Buttons - refers to designated points on the Macintosh screen which, when clicked with the mouse, moved the student to the next test item, recorded the student's answer to the current test item, or played a digitized sound.

Card - refers to one computer screen of information in the HyperCard development system. A card is very similar to a frame in traditional programmed instruction. Each card contained one test item and from two to four multiple-choice answers. The student was required to select the correct answer by clicking the mouse button.

Computer Lab - refers to the Music Computer Lab established in Room 319 of the Music Practice Building on the campus of Michigan State University. At the time of test administration the lab contained four Macintosh computers, two Apple II GS computers, one IBM computer, and various electronic keyboards.

Diagnostic test - refers to a test which evaluates the current achievement in the content area being tested. Diagnostic testing also implies the discovery of weaknesses or deficiencies in concepts necessary for completion of the test.

Digitized sound - refers to the capturing of actual musical sounds onto a computer disk. The process was accomplished with the MacRecorder from Farallon Computing, Inc. The MacRecorder hardware and software allow the user to record sounds through a microphone connected to the computer. The actua

the ir

digiti

evalu

prese

deve

exce

prog

of th

sim

gray

The

the

key

Ima

con

bu

con

the

ce

ac

actual sound was recorded on the computer disk and was played back through the internal speaker of the Macintosh computer. In the present study, digitized sounds were used to present aural examples for the students to evaluate.

Graphics - refers to any non-text items used in the computer presentation of the theory test. Graphics are an integral part of the HyperCard development system. This made possible the inclusion of small musical excerpts and other non-text enhancements.

Hardware - refers to all computer equipment but does not include the programs which run on the computer.

HyperCard - refers to the software used to create the computer version of the test. HyperCard allowed the presentation of each test item in a format similar to frames in a programmed text. It also provided a full complement of graphic design tools and accepted the inclusion of both sound and animation.

Macintosh - refers to the Apple computer used to administer the test. The Macintosh Plus and the Macintosh SE were used. The systems included the central processing unit, a black and white 9" monitor, a mouse, a keyboard, one or two floppy disk drives, a 20 megabyte hard disk, and an ImageWriter II printer.

Mouse - refers to a Macintosh computer input device. The user controlled the screen cursor by moving the mouse . Clicking the mouse button when the cursor was located in various parts of the monitor sent commands to the computer for processing.

Predictive test - refers to a test designed to indicate what may happen in the future. When a predictive test is valid it has been demonstrated that certain scores on the predictive test have a positive correlation with achievement in some criterion variable. In other words, a high score on the

predic
variab
an eva
taken

to do.

<u>Limit</u>

admi
conte
conte
at M
Mich
three
sign
minc
esse
espo
also
dev
dict

was
Un
stu
anc
be

predictive test indicates the likelihood of a high score on the criterion variable. The present study involved the development of a predictive test and an evaluation of the test's correlation with the criterion variables which were taken from various grades of freshman students in freshmen music theory.

Software - refers to the programs and files which tell computers what to do. For example, a word processing program is software.

Limitations

Several limitations were placed upon the development and administration of the test. Perhaps most important was the limitation on test content. Since expectations may vary greatly from school to school, the content of this test was developed in the context of music theory instruction at Michigan State University. The freshman music theory sequence at Michigan State University was quite traditional. Freshman students met three times each week for lectures covering musical concepts including key signatures, intervals, construction of major scales and the three forms of minor scales, triads, chord inversions, and modulations. There was essentially no introduction of 20th century theory methods such as those espoused by proponents of Schenker analysis or jazz studies. The students also met two days each week in a smaller aural skills lab. Here the students developed skills including a variety of exercises in sight singing, rhythmic dictation, melodic dictation, interval dictation, and chord dictation.

A second limitation involved the subjects used in the study. The study was limited to freshmen music theory students enrolling at Michigan State University during the Fall term of 1988. The assumption was that the students enrolling in the Fall term of 1988 would also enroll for the second and third term of the freshmen theory course and that these students would be comparable to future students at Michigan State University.

A

While th

revision

because

year. A

availabl

out of tl

F

done to

other h

theory

A limitation was also placed upon the subsequent revision of the test. While the test may profit from revision based upon results of this trial, revision of the test was not done as part of this study. This limitation is because of the fact that new subjects are only available once each academic year. An inherent problem with this limitation is the small size of the available sample and the great impact upon the study of students dropping out of the course.

Finally, the duration of the test was limited to 50 minutes. This was done to prevent undue boredom or loss of attention in the students. On the other hand, this time period allowed adequate time for a broad range of music theory topics.

A

areas: (

tests us

of predi

constru

<u>The Us</u>

I

achieve

testing.

for mu

place."

much a

(Lehm

for ma

measu

achiev

places

evalua

intere

might

CHAPTER 2 - REVIEW OF LITERATURE

A review of the literature relevant to this project encompasses five areas: (a) the usefulness of prediction, (b) the development or validation of tests useful in the prediction of some criterion variable, (c) the identification of predictive variables, (d) the use of computers in testing, and (e) the construction of a predictive musical test.

The Usefulness of Prediction

It is often difficult to distinguish between aptitude tests and achievement tests when reviewing the literature written about predictive testing. Typically, aptitude tests are designed to measure the "innate capacity for musical learning, even though no such learning may actually have taken place." Achievement tests, on the other hand, are "designed to measure how much a student has accomplished in music or in a particular phase of music" (Lehman, p. 8, 1968). Confusion arises when tests are used as predictive tools for making academic decisions or guidance suggestions. Does a predictive test measure what the student already knows thereby placing it in the realm of achievement tests or does it measure the student's capacity for learning which places it in the realm of aptitude tests? Very likely, various predictive tests evaluate both achievement and aptitude.

Lehman continued his discussion by outlining nine reasons for interest in musical testing:

1. Identification of talent. Tests provide early detection of talent which might go unrecognized.

inforn

untal

in sel

appro

stude

instru

not c

base

stude

stud

exce

distr

mus

than

mus

mea

occ

in a

2. Adaptation for individual differences. Tests give teachers the formation necessary to set challenging but attainable goals for the musically ntalented.

3. Educational guidance. Tests give the instructor information useful ı selection of the the proper musical instrument or in selection of ppropriately difficult academic coursework.

4. Vocational guidance. Tests provide objective information for the tudent considering a musical career.

5. Discovery of learning difficulties. Test results may allow the nstructor to detect and diagnose weaknesses. Even if these weaknesses are not correctable, the realization of their presence provides a better knowledge base for academic guidance.

6. Ability grouping. Tests may help the teacher place individual students with others of similar ability.

7. Assignment of instruments. Test results can be used to assign students to school-owned instruments when the number of applicants exceeds the number of instruments.

8. Studies of musical talent. Tests may reveal the extent and distribution of musical talent and the magnitude of individual differences.

9. Psychological studies. Tests can aid in research particularly when musical aptitude is used as a specific variable. (Lehman, 1968, p. 9)

Some of the previously listed rationales for testing are more substantial than others, but the list provides a strong foundation for further discussion of musical testing. Whybrew (1971) cited some practical reasons for measurement and evaluation in music. Musical evaluation is a frequent occurrence for teachers and students. Musicians frequently place themselves in adjudication situations where they expect objective evaluation. Music

teach
weak
in lig
stude
colleg
upon
admi
Why
have
direc
thos
iden
depa

First
goal
best
facto
cons
prec
exan
succ
hav
thar
fron
mu
per

teachers are constantly required to evaluate their students and diagnose weaknesses. Evaluation is a particularly necessary skill for college instructors in light of the responsibility placed upon them to properly advise incoming students and, when necessary, direct them into fields other than music. Each college or university makes its own decisions regarding any musical demands upon incoming students and there are very few apparent barriers to student admission designed to sift out untalented students. In another article, Whybrew stated that "recent emphases on accountability in education . . . have intensified the need for tools which would help music educators in directing their efforts more effectively and in demonstrating the results of those efforts more convincingly" (Whybrew, 1973, p. 9). One method of identifying untalented students is predictive testing within the music department itself.

Karma (1983) pointed out several pitfalls to avoid in predictive testing. First, the factors used in prediction should reflect the aims of the school. The goals of a particular course of study might not necessitate the selection of the best students. Second, effective prediction can only be achieved by using factors which actually affect success in the music study area under consideration. Therefore, careful research is necessary to identify valid predictors. Finally, successful prediction is the result of many predictors. For example, a musical aptitude test given to music students may not be a successful predictor by itself because the tested criterion, musical aptitude, will have a much smaller variance within the preselected group of music students than would be observed had the test been administered to college students from various majors. Other variables which might be combined with musical variables include intelligence, motivation, motor ability, and personality. Karma stressed one important fact, however: any variable used as a predic

limited tra

In li

is easy to

testing. D

test

shap

wit

con

axe

ans

Mr.

testing. T

the discon

about the

results. T

carpenter

accomplis

Developn

Th

upon the

previousl

tests dev

are the p

devoted

as a predictor must be selected on the basis of (a) its stability over time, (b) its limited trainability, and (c) its clear measurability.

In light of the problems associated with various prediction variables, it is easy to understand the division among writers when discussing musical testing. David Goslin of the Russell Sage Foundation stated,

> Attempting to predict future performance on the basis of test scores is much like trying to guess the ultimate size and shape of an oak tree by measuring a sapling in pitch darkness with a rubber band for a ruler, without taking into account the conditions of the soil, the amount of rainfall, or the woodsman's axe. The amazing thing is that sometimes we get the right answer. (Lehman, 1969, p. 19)

Mr. Gorlin has made a humorous point which holds true with music testing. The non-exact nature of some testing methods does not necessitate the discontinuance of testing, however. Rather, it requires new thinking about the goals of testing and the methods which produce the most useful results. Throughout history the perfection of adequate tools, whether for the carpenter or the researcher, has taken time and practice but has been accomplished in many areas.

<u>Development or Validation of Tests as Predictors</u>

This portion of the review of literature examines research focused upon the development of predictive tests and the validation and use of previously developed tests as predictors. The research described here includes tests developed for the specific purpose of prediction. The tests themselves are the predictive variable. A later section of the review examines research devoted to the identification of specific predictor variables other than a test.

One

Edwin Gorc

used with s

be used:

1. T

performanc

2. T

abilities of

3. T

4. T

5. T

Gorc

interested

however, e

value of th

Two

in testing c

test to 332

college stu

reasonably

that reliab

dissertatio

music stuc

student ev

Lee'

very usefu

Edwin Go

One of the most prominent individuals in musical test construction is lwin Gordon. His Musical Aptitude Profile (Gordon, 1965) has been widely ed with students of all ages. Gordon suggested five ways the test scores may used:

1. To encourage musically talented students to participate in music erformance organizations.

2. To adapt music instruction to meet the individual needs and bilities of students.

3. To formulate educational plans in music.

4. To evaluate the musical aptitude of groups of students.

5. To provide parents with objective information.

Gordon designed the test for younger school children and he was interested more in discerning aptitude than predicting success. Later studies, however, evaluated both the age of the individuals tested and the prediction value of the test .

Two studies examined the use of the Musical Aptitude Profile (MAP) in testing college students. In a 1967 project, Robert E. Lee administered the test to 332 college freshman music students to determine whether norms for college students could be established. He found that the test scores were reasonably reliable for college and university freshman music students and that reliable norms could be established. The study, based on Lee's doctoral dissertation, concluded that the scores of college and university freshman music students on the MAP were beneficial as one of many criteria used in student evaluation.

Lee's study and documentation of his research provide norms that are very useful to college educators particularly in light of a replicative study by Edwin Gordon in the same year (Gordon, 1967). Gordon's study involved

administrat
Lincolnwoo
of subjects
informatio
MAP, was
his test for
should be
adequate c
very mino
(Gordon, 1
The
music stu
include co
Lee's, that
his testing
training, h
have musi
in general
1972, p. 3
apparentl
conclude
nonmusic
student s
Fin
(1974). H
1947 by J
of Music

ministration of the MAP to freshmen at Rochester, Minnesota and ncolnwood, Illinois. No information is provided concerning the number subjects tested. Much of Gordon's article is a restatement of the formation Lee presented in his dissertation. Gordon, as the creator of the IAP, was willing to make a stronger positive statement regarding the use of is test for college students. He said, "The Musical Aptitude Profile can and hould be used as an educational diagnostic tool for the implementation of an adequate curriculum for college and university music students, and only to a very minor extent, if at all, should the battery be used as a 'talent' test" (Gordon, 1967, p. 40).

The previously mentioned studies dealt with the MAP and college music students. In a 1972 article, William T. Young expanded the research to include college and university nonmusic majors. His goal was similar to Lee's, that is, he wanted to establish norms for this particular target group. In his testing of 205 university students with little or no previous musical training, he discovered that nonmusic majors of the southern United States have musical aptitude "somewhat greater than that of high school students in general and lesser than that of midwestern college music majors" (Young, 1972, p. 390). Young's reference to midwestern college music majors was apparently a reflection upon the previously cited study by Lee. Young concluded that different norms must be established for music students and nonmusic students and that the MAP was a useful tool for diagnosing student strengths or weaknesses.

Finally, a study involving the MAP test was conducted by Schleuter (1974). He compared the Aliferis Music Achievement Test, first introduced in 1947 by James Aliferis, and two tests created by Edwin Gordon; the Iowa Tests of Music Literacy, Levels 5 and 6, and the MAP. The tests were administered to universit

diagnostic

developme

Test. Alife

coauthored

actually di

Schl

found that

subjects. F

provided t

since each

One

studies, or

other focu

1941). In

batteries c

defined co

harmony,

strength c

professior

applied p

success.

predictive

Taylor, th

likely to s

evaluatior

subseque

to university freshmen music majors in an attempt to determine the diagnostic strength of each test. James Aliferis was actively involved in test development and his important contributions include the Aliferis Freshman Test. Aliferis documented the construction of this test in an article coauthored with J. E. Stecklein (1953). The two Aliferis music tests are actually different tests with similar goals.

Schleuter acquired data for 150 subjects over a period of two years and found that each of the three tests provided useful information about the subjects. He concluded that the MAP test combined with an achievement test provided the most information. The choice of achievement tests is variable since each school has different objectives.

One early research study is especially interesting. It is actually two studies, one focused on the prediction of success in college music and the other focused on the prediction of success in the professional arena (Taylor, 1941). In the first study, Taylor evaluated the prediction strength of four batteries of music tests and one intelligence test upon college success. She defined college success as the ability to succeed in dictation, sight singing, harmony, and music history. The second study examined the predictive strength of the same four batteries of music tests upon success in the music profession. Using some of the same subjects tested in the first study, Taylor applied predetermined criteria to each subject to determine professional success. She concluded that none of the music test batteries have sufficient predictive power to be used by themselves in student guidance. According to Taylor, the student who is successful in dictation and sight singing is most likely to succeed professionally. A final conclusion stated that the evaluations by a student's instructors are very reliable indices of the student's subsequent success in professional music. Although many of the more

recently de
study prov
researcher
using the r

Ther
1960s and
pertain to
variables u
found that
psychomo
study, the
intelligenc
aptitude te
small num
tentative.

Go
1984 and
discrimina
validity a

In
1970. In
test batte
This inclu
out, on th
according

In
documen

recently developed tests were not available when Taylor did her research, this study provides insight into the predictive strength of the tests used. A researcher could produce an interesting study by replicating Taylor's research using the modern music tests currently available.

There was a marked increase in research of music testing in the late 1960s and early 1970s. Several studies deserve mention here since they pertain to predictive testing. Hufstader (1974) undertook the identification of variables useful as predictors of success in beginning instrumental music. He found that musical aptitude, academic achievement, intelligence, and psychomotor skills all contributed to the prediction of success. In another study, the opposite conclusion was reached (Gordon, 1968). That is, intelligence and achievement tests do not enhance the predictive power of aptitude tests. The strength of Gordon's conclusion was weakened by the small number of subjects and the author's admission that the findings were tentative.

Gordon continues to be active in testing, especially with children. In 1984 and 1986, Gordon completed longitudinal studies of his auditory discrimination and timbre preference tests. These studies in predictive validity are of general interest.

In the area of college testing, two important studies were completed in 1970. In an examination of test content, Whellams (1970) found that aptitude test batteries should include non-musical tests as well as aural-musical tests. This inclusion increases the predictive strength of the aptitude test. He points out, on the other hand, that the types of non-musical tests included vary according to the social and educational background of the subjects.

In a study with specific impact on the research reported in this document, Ernest (1970) found that the best single predictor of college grade

point (r=.4
addition o
enhance th
In a
Ohio State
remedial c
indicated t
in the rem
training c
emphasizi
present re
freshmen
Sch
Seashore
After test
the two t
much mo
Seashore
Kwalwas
Ba
test in m
to measu
discrimin
sections
single m
A
tests in

point (r=.43) and music grade point (r=.44) was high school rank. The addition of nonmusical aptitude and achievement tests did not significantly enhance the predictive ability of high school rank used alone.

In a 1983 study, Arenson sought to validate the music portions of the Ohio State University Entrance Battery as a predictor of success in the two remedial courses offered at the University of Delaware. His findings indicated that the OSU theory combined score was a good predictor of grades in the remedial course emphasizing cognitive knowledge. The OSU ear-training combined score was a good predictor of grades in the remedial course emphasizing ear-training and listening. This study is very pertinent to the present research study since it involves the prediction of academic success of freshmen students in theory.

Schmitz (1956) investigated the prognostic value of the revised Seashore tests and the Kwalwasser-Ruch Test of Musical Accomplishment. After testing 582 students who were administered various combinations of the two tests, Schmitz found that grades below the mean were predicted with much more accuracy than grades above the mean. The B form of the Seashore tests appeared to be the strongest single predictor while the Kwalwasser-Ruch Test was not a strong predictor.

Ball completed a study involving the construction of a college entrance test in music in 1964. He constructed a battery of thirteen musical ability tests to measure rhythmic, melodic, and harmonic abilities as well as interval discrimination, chordal analysis, and memory. His research indicated the sections involving memory, interval discrimination, and discrimination of single music elements were the best predictors.

A study by Perry (1965) examined the predictive proficiency of selective tests in music theory administered individually and in groups. His goal was to determi

music theo

strength co

students.

Proficiency

After adm

of the test

greater th

The Identi

A r

in predict

have been

research i

from plac

requirem

receive le

concepts

Or

school gr

student's

1970; Ch

rank we

exceptio

A

with a s

correlat

1973 stu

to determine which tests were good predictors of performance in college music theory courses. Tests that provided a significant level of predictive strength could be used in guiding, counseling, placing, selecting, or grouping students. The predictor tests were administered prior to the start of classes. Proficiency tests were given after the first semester of theory was completed. After administering the test to 91 freshmen students, Perry found that seven of the tests under investigation were significant predictors with correlations greater than r=.60.

The Identification of Predictive Variables

A number of research studies have sought to identify variables useful in prediction of academic success in various musical areas. Many variables have been examined as predictors. One of the problems with this type of research is that the variables useful in prediction of success may vary greatly from place to place. This is typically the result of non-standardization of requirements. Concepts which receive great attention at one location may receive less emphasis at another. Therefore, any test emphasizing certain concepts is likely to be more useful in one school than in another.

One of the most important prediction variables identified thus far is school grade point average. A weaker form of this variable is found in a student's class rank. Several studies (Horst, 1959; Turrentine, 1965; Ernest, 1970; Chadwick, 1976; Hedden, 1982) found that grade point average or class rank were significant predictor variables. Each of these studies, with the exception of Hedden, focused upon college level testing.

Another strong predictor variable, intelligence, is usually measured with a standardized intelligence test. Neely (1965) found a positive correlation between intelligence and notational ability in ear-training. A 1973 study showed that musicality and intelligence could function as predictors

placed inte

musical tra

combinatio

viable pre

Rey

personality

Krueger fo

predictors

factor of t

other vari

perspectiv

the perspe

suggested

directing

whether t

He found

school an

of partici

Tw

examine

that stro

harmonic

echo-play

accompa

In

instrume

predictors of choral achievement (Helwig & Thomas, 1973). Another study placed intelligence in a long list of variables which included aptitude tests, musical training, personality, age, sex, race, home environment, and various combinations (Webber, 1976). Each of these studies found intelligence to be a viable prediction variable.

Reynold Krueger did rather extensive research into the variable of personality as a predictor of teaching success. In two studies (1972 &1976), Krueger found that personality and motivation were very powerful predictors of teacher success. The power of these variables, however, is a factor of the measurement instrument used to gather data and the control of other variables. Motivational variables have also been studied from the perspective of success at high school band directing (Caimi, 1984) and from the perspective of college ensemble participants (Mountford, 1982). Caimi suggested that insufficient numbers of motivational variables exist in band directing tasks to warrant prediction of success. Mountford examined whether there are variables useful in predicting college band participation. He found that variables such as extracurricular use of instrument in high school and nonselection of rock as a favorite style were significant predictors of participation.

Two studies do not fit neatly into a variable category since they examine unusual predictor variables. One study (Humphreys, 1986) found that strong ability to echo-play a melodic segment indicated success at harmonic audiation and performance. Humphreys suggested that training in echo-playing may enhance a student's ability to play implied harmonic accompaniments.

In a 1981 study, Brand and Burnsed researched whether the number of instruments played, ensemble experience, GPA in music theory, GPA in

sightsingin
could func
the examin
that error
instrumen
instrumen
One
with intel
attainmen
intelligenc
listening a
was best
best predi
used as a
A 1
students
attempt t
pitch adj
of intona
the hypo
could sig
Se
there is a
measurin
also doc
succeed.
interest i

sightsinging and ear training, or years of private instrumental instruction could function as predictors of error detection ability. Unfortunately, none of the examined variables proved to be effective predictors which may indicate that error detection skill is not developed in the same fashion as other instrumental music abilities or it may indicate that the measurement instrument was not sufficiently reliable to demonstrate a correlation.

One important study (Young, 1969) combined the Gordon aptitude test with intelligence and academic achievement tests to predict musical attainment. Young found that the MAP and either an achievement test or intelligence test were the best predictors of success in performance and listening areas of music. Conversely, success in the academic areas of music was best predicted by an intelligence test. Overall achievement in music was best predicted by the three types of tests (aptitude, intelligence, achievement) used as a group.

A 1982 study by Chevallard used 77 undergraduate and graduate students in applied voice, woodwind, and brasswind instruction in an attempt to determine whether pitch memory, pitch discrimination ability, pitch adjustment ability, or pitch steadiness ability could be used as predictors of intonational performance. However, all research studies do not produce the hypothesized conclusion and Chevallard found that none of the variables could significantly strengthen the prediction of intonational performance.

Several conclusions are drawn from the research cited above. First, there is a continuing interest in musical testing. Researchers are desirous of measuring the characteristics which mold musical ability. The studies cited also document the interest of researchers in predicting which students will succeed. This interest spreads across all age groups. Finally, motivated by an interest in predicting student success, researchers have tested a broad range of

musical va
this same
<u>The Use o</u>
The
outset. In
topic of co
uses of co
outside th
includes s
This speci
Much of
who use
choices.
conceptio
counselir
In
out that
informat
be provi
simplifie
inputtin
Some in
human
impleme
Ir
better o
college

musical variables to identify those which function as strong predictors. It is this same motivation which propels the study reported in this document.

The Use of Computers in Testing

The parameters of this section of the review must be defined at the outset. In the past fifteen years a large body of articles has appeared on the topic of computer-assisted instruction, computer-assisted learning, and the uses of computers in education. For the most part, this body of research falls outside the scope of this review. The literature included in this section includes studies directed toward examination of computer uses in testing. This specific area is still in its infancy and is especially undeveloped in music. Much of the research in this field is aimed toward school guidance counselors who use the computer as a tool to direct students in academic and career choices. This has bearing upon the present research since it is hoped that the conception of a theory test will lead to the development of an academic counseling tool.

In a study on computers in counseling, Eberly and Cech (1986) pointed out that "computer technology permits presentation of more precise information without oversight or observer bias at a greater speed than could be provided by a counselor" (p. 18). They go on to state that computer usage simplifies the collection of data and increases the privacy of the individual inputting the data but these advantages are not without negative aspects. Some individuals may view the computer as an inadequate replacement for a human teacher. Thus, they are less likely to cooperate with attempts to implement the new technology or to see the computer as a benefit.

In the area of testing, the prime question is whether computer testing is better or worse than standard pencil and paper testing. A study, using 72 college students sought to answer this question (Fletcher & Collins, 1986).

The study

of a test w

pencil test

computer

following

1.

2.

3.

4.

The

They incl

1.

2.

3.

All

study in

determin

alleviatio

Th

professio

are more

computer

Th

It is now

Examina

under de

response

The study found that the mean scores of students taking a computer version of a test were roughly equivalent to the scores of students taking paper and pencil tests. The study also demonstrated that most students preferred the computer version of a test over paper and pencil test versions for the following reasons:

1. Computers can provide immediate scoring.
2. Computers can provide immediate feedback on incorrect answers.
3. Computers are more convenient, straight forward and easy to use.
4. Computer tests are completed more quickly than written tests.

The students also identified some disadvantages to computer testing. They included:

1. Inability to review all responses.
2. Inability to make changes to responses.
3. Inability to skip questions and return to them later.

All these disadvantages were a product of the test used in the research study in which these students participated and were design considerations determined by the test developer. Current technology allows for the alleviation of each of the listed disadvantages.

The results of the previously cited study appear to have support in the professional arena as well. A recent study suggested that adolescent students are more willing to input information into a computer since they view the computer as less threatening than an adult (Millstein, 1987).

The development of computer testing is moving ahead at a rapid pace. It is now possible for students to take practice forms of the Graduate Records Examination (GRE) with microcomputers (McArthur & Choppin, 1984). Also under development are systems which will diagnose patterns of error in responses to multiple choice questions. One of the most recent developments

is adaptive
answer to
informatio
no other q
designed t
the compu

San
As has alr
The numb
for this re
cited belo

1.
traditiona

2.
needs.

3.
response

4.
time mus

5.

6.
response

T
Howeve
Knowle
importa
fear mig

is adaptive testing. An adaptive test varies with each response. A correct answer to a certain question supposedly demonstrates mastery of all the information necessary for that answer. The computer "adapts" the test so that no other questions covering that material are asked. Elaborate systems can be designed that remember which errors the student has made in the past and the computer can provide constant remedial help with problems.

Sampson (1983) pointed out the potential benefits of computer testing. As has already been stated, there is a positive response to computer testing. The number of advantages inherent to computer testing may be responsible for this response. A partial list of advantages gained by computer testing is cited below.

1. Computer testing has proven to be at least as cost effective as traditional testing.

2. Adaptive testing allows for specialized attention to individual needs.

3. The computer can generate a wealth of data along with test responses.

4. Since the computer handles many of the administrative tasks less me must be spent by staff persons.

5. Administration and scoring of tests is more flexible and efficient.

6. Student error rates are decreased. That is, errors such as placing ponses in the wrong number are eliminated.

These advantages were reiterated by Meier and Geiger (1986). ever, Sampson lists some problems along with the advantages. ledgeable persons can tamper with records making security an tant issue. Some individuals have a fear of using computers and this ght be reflected in their performance. Although these are very real problems,

advantage

disadvanta

Tur

The comp

Tests can

the test it

student w

Bej

fact, he w

computer

decrease

pencil tes

within co

within ir

Compute

the comp

tradition

T

perform

criterion

behavio

style cla

measure

student

finding

problems, they are surmountable and do not necessarily diminish the advantages of computer testing. One must accept trade-offs of advantages and disadvantages with any form of testing.

Turner (1987) added another advantage to those listed by Sampson. The computer allows the test administrator to create large banks of test items. Tests can then be generated from these banks of items. If sufficient analysis of the test items is completed, it is possible to generate a different test for each student while maintaining equivalent item difficulty.

Bejar (1984) agreed with the stated advantages of computer testing. In fact, he went one step further and pointed out that in some instances a computer test is preferable to the traditional method. Some variables which decrease precision of score assessment cannot be controlled in a paper and pencil test. Typical scoring of paper and pencil tests focuses on variance within correct responses. Computer scoring allows analysis of variance within incorrect responses as well as variance within correct responses. Computer scoring also provides complete error control during scoring and the computer can generate information which is not readily available with raditional methods of scoring.

Two important studies of computer testing in music have been rformed. In 1972 Radocy evaluated the viability of using computers for terion-referenced testing of nonperformance music behaviors. The aviors examined by Radocy included dictation, interval recognition, and classification. Radocy developed a test based on behavioral objectives to ure competency in the stated objectives. The test was administered to 32 nts by the computer and 28 students by conventional methods. Radocy's gs have tremendous impact upon the present study.

1. I

the constr

certain nc

2.

success of

made reg

3.

nonperfor

conventio

cited abo

Mu

and Sims

music tra

the subje

keyboarc

subject.

also reco

subject t

interest

adminis

Constru

W

tools sh

typical I

facts to

known

new pr

1. Present skills (in 1972), techniques, and equipment are adequate for the construction of a workable computerized criterion-referenced test of certain nonperformance musical behaviors.

2. Rank order of items, in terms of item difficulty, is critical to the success of an incremental programing strategy wherein assumptions are to be made regarding responses to nonadministered items.

3. The computerized criterion-referenced test of certain nonperformance musical behaviors is not at present equivalent to a conventional paper-and-pencil version of the test. (The more recent studies cited above may refute this finding.)

Music preference has also been the object of computerization. Gregory and Sims (1987) developed a computer program to present nine four-voice music transcriptions to the subject in random order. The computer allowed the subject to change the music selection at any time by pressing a key on the keyboard. The computer then recorded the elapsed listening time for each subject. In a second study with the same hardware and software the computer also recorded the subject's like or dislike of each music selection when the subject touched the appropriate box on the screen. This study is of special interest since it demonstrates the use of computers as unattended test administrators and scorers.

Construction of a Predictive Musical Test

Wedman and Stefanich (1984) stated that computer based assessment tools should test concepts, principles, and procedures as well as facts. In a typical learning sequence the student begins by committing a particular set of facts to memory. Second, the student learns to restate and interpret the known facts. Finally, the student is able to apply and use the facts to solve new problems created through various situations. It follows, then, that the assessmen

respond a

suggest th

1. I

evaluatio

2.

examples

3.

principles

the learn

4.

perform

procedur

Al

tremend

Markle (

test desi

O

predict

howeve

stated t

that is e

(p. 64).

(

the dev

selectin

test de

assessment tools must incorporate items which will encourage the student to respond at higher levels than recitation of facts. Wedman and Stefanich suggest that successful computer assessment requires the following:

1. Determine the type or types of content to be included in the evaluation.

2. For conceptual content, test items should have the learner select examples from non-examples for each of the concepts included.

3. For principle content, test items should have the learner apply the principles in ways consistent with how the principle will be applied outside the learning situation.

4. For procedural content, test items should require the learner to perform the procedure under conditions similar to those in which the procedure will be performed away from the learning situation. (p. 27-28).

All of these guidelines will not apply to the present study but they are a tremendous help in channeling development ideas. Other documents by Markle (1969) and Bloom and Peters (1961) presented helpful information on test design.

One of the desired outcomes of the proposed study is the ability to predict success in music theory of incoming college students. The test itself, however, will be a diagnostic, criterion-referenced test. Willingham (1974) stated that criterion-referenced tests "should provide diagnostic information that is especially relevant to placing students and monitoring their progress" (p. 64).

Colwell (1970) suggested several characteristics which are important to the development process. He makes the suggestions as guidelines for selecting an appropriate test. However, they are necessary considerations in test development. Factors to evaluate are time, difficulties in administration, cost versus

assessed, c

evaluation

desired va

selecting a

(Gronlund

Sev

including

Content v

importan

ability of

present s

related v

under co

the test b

reliabilit

R

assessm

In other

score hi

measur

method

two sub

include

correla

metho

takes t

cost versus value received, and difficulties in scoring. Once these factors are assessed, one of the most important areas of test development is the evaluation of validity. Validity evaluates whether the test measures the desired variables. Colwell labels this the most important consideration when selecting a test. Many other writers stress the need for strong validity as well (Gronlund, 1965; Berk, 1980; Swezey, 1981; Karma, 1982).

Several types of validity are often imposed upon prediction tests including content validity, criterion-related validity, and construct validity. Content validity and criterion-related validity are perhaps the most important measures for this study. Content validity is an assessment of the ability of the test items to evaluate the given content areas. Often, as in the present study, content validity is determined by a panel of experts. Criterion-related validity is important since it analyzes whether the scores on the test under construction correlate positively with another criterion. Sometimes the test becomes its own criterion measure which leads to the discussion of reliability.

Reliability is a major concern in prediction testing. Reliability is the assessment of the internal consistency of the test or the stability of test scores. In other words, if a student scores high on one set of items, they should also score high on an equivalent set of items. Internal consistency can be measured a number of ways with the odd-even formula and the test-retest method being the most popular. The odd-even formula divides the test into two subtests. One subtest includes the even items while the other subtest includes the odd items. The two scores obtained from the subtests can be correlated to determine the strength of internal consistency. The test-retest method is similar except that the entire test provides both scores. The student takes the test one time and then retakes the test after a predetermined period of time ha

forget test

knowledge

predictive

previous p

Per

a book by

comprehe

evaluation

supportin

with testi

formats a

statistical

<u>Conclusi</u>

Th

in music

compute

tools in t

helpful t

of time has elapsed. The time period is long enough to allow the student to forget test content but short enough to discourage large increases in knowledge. In a sense it is the reliability of a test that determines that test's predictive strength. With the exception of Karma, the authors listed in the previous paragraph all stress the importance of test reliability.

Perhaps the most important musical source on test construction is the a book by Boyle and Radocy (1987). This text is an exceptionally comprehensive source with specific relativity to music measurement and evaluation. Boyle and Radocy took particular care to discuss the reasons supporting measurement of music experiences, the difficulties associated with testing, the steps necessary for proper test construction, various test formats available for evaluating specific musical experiences, and the statistical analyses which are most useful in evaluating the gathered data.

Conclusion

The cited literature demonstrates the usefulness of testing, the interest in music testing, the variables tested in the past, and the viability of the computer as a tool for test administration. The need for reliable diagnostic tools in the area of music is great. The development of such a test could be helpful to college instructors as well as students.

The

facilitate t

test const

documen

create the

administ

Th

content a

State Un

educatio

the instr

Michigar

students

knowled

the exte

test scor

successf

the curr

T

was ex

develo

intervi

list the

CHAPTER 3 - DEVELOPMENT OF THE TEST

The development of any test includes a number of steps designed to facilitate the creation of a useful measurement tool. Several references for test construction are located in the review of literature included with this document. The following discussion outlines the procedural steps taken to create the theory test used in this research study. The procedures for test administration are presented in Chapter 4.

The first step taken in the creation of this test was the development of content areas with the instructor of the first-year theory sequence at Michigan State University. The inclusion and sequencing of material in many educational settings are determined by the assumptions and expectations of the instructor. The course content of the first-year theory sequence at Michigan State University is based upon the assumption and expectation that students enrolled in the class have already achieved a certain level of musical knowledge. A student could, through pre-enrollment testing, demonstrate the extent of achievement of the expected musical concepts. The generated test score could provide an indicator of the student's probability of successfully completing the first-year theory sequence. This is the premise for the current study.

The first goal was to determine the musical content areas the student was expected to know upon enrolling in the first-year theory sequence. The development of test item content areas was started through a personal interview with the first-year theory instructor. The instructor was asked to list the specific musical knowledge he expected each student to have mastered p

the most s

opportuni

to the mea

In t

Harder's

who perfo

chosen be

preparati

in Tonal

first-year

indicates

enrolling

additiona

Theory t

final list

Th

formulat

reflected

Gronlun

as a gen

since it

A

purpose

behavio

objectiv

content

mastered prior to taking the course. A face-to-face discussion was chosen as the most suitable means of gathering this data since it allowed me the opportunity to ask questions so that the instructor and I were in agreement as to the meaning of each content area listed.

In the past, the theory area at Michigan State University used Paul Harder's Basic Materials in Music Theory (1986) as a remedial text for students who performed poorly in the freshman theory sequence. This text was chosen because it was written to guide students through basic music theory in preparation for extended study in Harder's larger work, Harmonic Materials in Tonal Music : A Programed Course (1985) which is used as the text for the first-year theory sequence. The utilization of Basic Materials in Music Theory indicates that it addresses many of the content areas expected of students enrolling in the first-year theory sequence. Based on this assumption, additional content areas were selected from the Basic Materials in Music Theory text to supplement the list already compiled by the instructor. The final list of content areas was then approved by the instructor.

The next goal after compilation of relevant content areas was the formulation and clarification of one or more behavioral objectives which reflected each content area. Two references were helpful in this process. Gronlund's (1985) brief but specific text on developing objectives was valuable as a general reference. Boyle's (1974) book was more applicable to this study since it dealt specifically with developing test objectives in music applications.

A panel of theory experts was selected at this point in the study. The purpose of this panel was to (a) evaluate the appropriateness of the behavioral objectives, (b) suggest possible revisions to the behavioral objectives, (c) suggest appropriate proportions for the test to sample from each content area, (d) evaluate whether test items properly test the behavioral

objectives,
important
developm
concernin
Th
Ruggiero,
had num
Dr
Michigan
Michigan
Fellowsh
he has ta
Lansing
of the La
and viol
biograpl
lectured
Univers
D
Conserv
Current
has tau
founde
Music
univers
used a
compo

objectives, and e) suggest possible revisions to test items. The panel was important in establishing the content validity of the test at all stages of development. The researcher retained control of all final decisions concerning design revisions to the test.

The panel of experts consisted of Dr. Theodore Johnson, Dr. Charles Ruggiero, and Dr. Corliss R. Arnold. Each of the individuals on the panel has had numerous and varied experiences in the field of music theory.

Dr. Theodore Johnson, Professor and former Chair of Music Theory at Michigan State University, earned a doctorate from the University of Michigan, where he was awarded the Stanley Medal and a Rackham Fellowship. A U. S. Army veteran (Korea) and Fulbright Scholar (Munich), he has taught at the University of Kansas, has served as Concertmaster of the Lansing and Grand Rapids Symphony Orchestras as well as Principal Violist of the Lansing Symphony Orchestra, and has performed widely as a violin and viola soloist and chamber musician. Dr. Johnson--who is a subject of biographical record in several editions of Who's Who in America--has also lectured at various universities and authored two books, published by University Press of America (1982 & 1986).

Dr. Charles Ruggiero holds degrees from the New England Conservatory and Michigan State University (Ph.D. in composition, 1979). Currently, Dr. Ruggiero is a professor at Michigan State University where he has taught music theory and composition since 1973. Dr. Ruggiero is the founder and current co-director of Michigan State University's Computer Music Studio and he has initiated several computer music courses at the university. His Set Analysis Programs computer software published in 1986 is used at universities throughout the United States. Dr. Ruggiero is a noted composer with such pieces as Three Blues for Saxophone Quartet (1982), Interplay

1988, Dr.

School of

Dr.

Michigan

study in

by the Ar

Compreh

choral co

He has ta

Th

submitte

evaluatio

of the ob

the conc

their ori

were inc

maintair

included

Unless c

aural ite

answers

1.

2.

3

level or

4

Interplay (1989), and Dances and Other Movements (1989) to his credit. In 1988, Dr. Ruggiero was appointed Chair of the Michigan State University School of Music's Theory Area.

Dr. Corliss R. Arnold is Professor of Pipe Organ and Music Theory at Michigan State University. In 1956-57 he held a Fulbright Fellowship for study in France. Dr. Arnold has earned the two highest certificates awarded by the American Guild of Organists. One of his books, Organ Literature: A Comprehensive Survey (1984), is in its second edition. Dr. Arnold is also a choral conductor and composer of works for organ, chorus, and instruments. He has taught at Michigan State University since 1959.

The behavioral objectives created from the original content areas were submitted to the panel. The main goal of the panel review was the evaluation of the behavioral objectives. Each panel member was given a list of the objectives and asked to approve, modify, or reject each objective. At the conclusion of the panel review, all objectives were approved, either in their original form or with revisions. All revisions suggested by the panel were incorporated in a modified list of objectives with special care taken to maintain the behavioral nature of each objective. The final list of objectives included seven content areas with several objectives under each content area. Unless otherwise indicated, all stimuli were written, not aural. To respond to aural items, the subject was asked to select a response from a set of visual answers on the screen. The objectives for pitch notation were:

1. Identify bass and treble clefs by their signs.

2. Name notes positioned on bass and treble clefs.

3. Rewrite notes from one clef to another while maintaining pitch level or transposing by an octave.

4. Identify half steps and whole steps in a melodic example.

5.

sharp, an

6.

7.

Ti

1.

2.

3.

4.

based u

Th

1.

shortest

2.

shortest

3.

stems.

4

5

I

1

2

3

interva

4

I

1

5. Recognize accidental signs including sharp, flat, natural, double-sharp, and double-flat.

6. Select the enharmonic equivalent of a given note.

7. Recognize examples of chromatic scales.

Time notation objectives included:

1. Identify measures with correct notation based on a given meter.

2. Distinguish between examples of simple and compound meter.

3. Match a triplet with another note grouping of equal duration.

4. Identify the tempo of a specific note type (eighth notes, for example) based upon a given metronome marking.

The area of notes and rests included the following objectives:

1. Arrange notes in order from longest duration to shortest duration or shortest duration to longest duration.

2. Arrange rests in order from longest duration to shortest duration or shortest duration to longest duration.

3. Identify correct or incorrect placements of note flags, beams and stems.

4. Recognize the effect of the dot upon note and rest durations.

5. Recognize the effect of ties upon given note durations.

Interval objectives were:

1. Identify harmonic and melodic intervals from aural examples.

2. Identify compound intervals.

3. Recognize major, minor, perfect, augmented, and diminished intervals.

4. Correctly identify the inversion of an interval.

For the scales portion of the test the objectives were:

1. Identify correct examples of major, minor, and other modal scales.

2.

3.

scale.

4.

scales.

5.

6.

Th

1.

2.

3.

4.

F

1

2

3

4

aural e

the im

develo

assign

and av

25%, (

triads

the lis

areas

2. Distinguish between examples of chromatic and diatonic scales.

3. Identify the series of half steps and whole steps in a major or minor scale.

4. Recognize examples of natural, melodic, and harmonic minor scales.

5. Recognize correct labels for each scale degree.

6. Identify major, minor, and chromatic scales from aural examples.

The sixth area, key signatures, included the following objectives:

1. Select the key indicated by a given key signature.

2. Identify the parallel and relative minor key of a given major key.

3. Identify the parallel and relative major key of a given minor key.

4. Identify through inference the key of a given melody.

Finally, objectives in the area of triads were:

1. Build a triad around a given root, third, or fifth.

2. Identify the scales which allow the building of a given triad.

3. Label first and second inversions of triads.

4. Identify major, minor, augmented, and diminished triads from aural examples.

After reviewing the final list of objectives, the panel was asked to rank the importance of the objectives so that a Table of Specifications could be developed. The panel chose to rank the broad content area descriptions assigned by the researcher rather than individual objectives. The rank order and average percentage of importance assigned by the panel were (a) scales - 25%, (b) pitch notation - 20%, (c) notes and rests - 15%, (d) intervals - 15%, (e) triads - 15%, (f) key signatures - 10%, and (g) time notation - 5%. To achieve the listed percentages, the panel was first asked to rank the seven content areas by perceived importance. All panel members ranked the items as they are listed

percentag

and again

balance o

objective

small dif

was imp

category.

Us

Specifica

was use

percenta

produce

are listed above. The panel was asked to weight each content area with a percentage rounded to the nearest factor of five. The weights were averaged and again rounded to the nearest factor of five to achieve a symmetrical balance of 100%. It was the view of the panel that assigning weight to each objective would not provide a substantial increase in information due to the small differences between some objectives. The panel also suggested that it was important to obtain close approximations of percentages within each category.

Using the percentages listed above as a guide, the Table of Specifications shown in Table 1 was generated. The Table of Specifications was used to develop a group of 90 objective test items. Some content area percentages were modified slightly with the agreement of the panel to produce an even distribution of questions within content areas.

Table 1

Number o
Three Typ

Weight
27%
20%
13%
13%
13%
8%
6%

T
Knowle
which h
learned
informa

A
develop
based o
Multip
was po
objecti
was als

Table 1

Number of Test Items in Each Content Area of the Colman Test Divided Into Three Types of Mastery

		Learning Level		
Weight	Content Area	Knowledge	Understanding	Application
27%	Scales	10	10	4
20%	Pitch notation	7	6	5
13%	Notes and rests	6	4	2
13%	Intervals	5	3	4
13%	Triads	1	5	6
8%	Key signatures	6	0	1
6%	Time notation	2	2	1

The three types of mastery listed in Table 1 were defined as follows. Knowledge is the ability of an individual to recall facts or other information which has been learned. Understanding is the ability to interpret previously learned information. Application is the ability to take previously learned information and apply it to some new problem or circumstance.

A decision was made to employ multiple-choice items when developing the test. This form of test item was chosen for several reasons based on Marshall and Hales' (1971) evaluation of various test formats. Multiple choice items are strong in the areas of flexibility and versatility. It was possible to formulate multiple-choice test items capable of measuring objectives at all levels of the cognitive domain. A multiple-choice format was also widely applicable since it was adaptable to all subject areas and grade levels. C

since mar

Finally, M

anyone, e

accurately

created b

particula

of test ite

necessary

M

testing.

kı

it

th

M

of an ex

underst

second

since th

study.

experts

test ite

Each te

fostere

clarific

levels. Coverage of a broad spectrum of instructional objectives was possible since many test items could be answered in a normal examination period. Finally, Marshall and Hales suggested that multiple-choice tests allow anyone, even those unfamiliar with the course material, to rapidly, accurately, and objectively score the test without the intrusion of any bias created by past performance, attitudes, appearance, or other factors. This was particularly important since the present research involved computer scoring of test items. The computer could not make the kinds of judgement decisions necessary for accurate scoring of essay type tests.

Marshall and Hales also included some weaknesses of multiple-choice testing.

> In order to develop good items, the writer must have a thorough knowledge of the course content, an awareness of the methodology of item writing, skill in the use of language, and a thorough knowledge of the level of development of [the students involved in the test]. (p. 93)

Multiple-choice tests also fall short when the goal is the measurement of an examinee's ability to organize materials or clearly express an understanding of the required material in an acceptable writing style. This second objection to multiple-choice testing did not impact the present study since the testing of these abilities was not important to the success of the study.

The 90 completed test items were then submitted to the panel of experts for evaluation of content validity. The panel was asked to critique the test items for clarity and successful measurement of the desired objective. Each test item was listed together with the behavioral objective which fostered it to facilitate the evaluation process. The panel suggested various clarifications and modifications which were subsequently incorporated into the final v

into a con

was comp

Th

vehicle fo

were acc

response

chance t

The choi

a new c

Univers

multiple

speed o

T

comput

publish

present

the scr

answer

animat

frames

item n

respon

Screen

the sc

clickin

the final version of the test (see Appendix D). The process of placing the test into a computerized format was begun once the modification of test items was completed.

The Apple Macintosh computer was selected as the test administration vehicle for this project because of its ease of use. All responses to test items were accomplished by clicking the mouse button to select the desired response. The goal of this design was to alleviate or at least greatly lessen the chance that fear of using a computer would adversely affect test performance. The choice of the Macintosh computer was also influenced by the existence of a new computer music lab in the School of Music at Michigan State University. The installation of several Macintosh computers in this lab made multiple administrations of the test possible, which greatly increased the speed of data collection.

The implementation of the music theory test on the Macintosh computer was accomplished using the HyperCard (Atkinson, 1988) program published by Apple Computer. With HyperCard, the 90 test items were presented in a linear fashion. Each test item was displayed individually on the screen in the form of a "card" and the user responded by selecting an answer with the mouse. HyperCard also allowed the use of sound, animation, and graphics to enhance test presentation.

The creation of the test involved several programming steps. First, 90 frames were created to present the test items. Each frame included the test item number, the test item, the box which displayed the letter of the selected response, and up to four response choices represented by screen buttons. Screen buttons on the Macintosh computer were visual representations on the screen of real buttons. The user passed information to the computer by clicking the mouse on one of the buttons, specifically, the button which

displayed
also conta
button in
up the in
test func
marker f
selected,
compute
impleme
prohibit
answer
moved
progres
at the e
allowe
require
respo
prohib
blank.

visua
and r
Maci
Profe
Profe
trans
befo

displayed the selected response. Along with the answer buttons the frames also contained several buttons which affected test presentation. The "Help" button in the upper right corner of each frame enabled the examinee to call up the instruction screens presented at the beginning of the test if any of the test functions became unclear. The "Come Back Later" button functioned as a marker for questions the examinee wanted to answer later. If this button was selected, the examinee could complete the rest of the test items before the computer automatically returned to the marked questions. The implementation of this feature required careful test item sequencing to prohibit the possibility that later questions might provide information to answer previous questions. The "Arrow" button in the lower right corner moved the examinee from one question to the next. Only a forward progression was allowed with the exception of a return to marked questions at the end of the test. One of two options was required before the computer allowed the examinee to proceed to the next question. The examinee was required to check the "Come Back Later" button or select one of the possible responses which was then recorded in the response box. This feature prohibited examinees from accidently or intentionally leaving test items blank.

The development of the test frames was followed by the creation of all visual examples used in the test items or responses, including notes, symbols, and musical excerpts. All examples were generated with the help of the Macintosh computer. First, the musical notation was programmed into Professional Composer (1988), a computer music notation program. Professional Composer generated a graphic form of the notation which was transferred to SuperPaint (Gay, 1989), a graphics program for final editing before placement into the desired test frame. This phase of test development also inclu

and sour

O

the deve

intervals

playing

compute

did not

additio

mechan

aural ex

comput

hardwa

hardwa

recordi

sounds

compu

compu

called

examp

were a

the M

speak

test.

theor

the co

also included the creation of instruction screens complete with animation and sound.

Once the visual portion of the test was completed, attention turned to the development of the aural portion. The test required aural examples of intervals, scales, and triads. The original test design called for the capability of playing these aural examples through a musical keyboard attached to the computer. However, this plan was altered because the HyperCard program did not offer ready access to transmitting the sound information and the addition of further equipment greatly increased the possibility that mechanical failure could adversely affect test administration. Instead, all aural examples were reproduced from actual digitized sounds stored on the computer disk. This was accomplished with the MacRecorder (Capps, 1988) hardware and software from Farillon Computing. The MacRecorder hardware attached to the back of the Macintosh computer and functioned as a recording microphone. The software portion of the package captured the sounds coming through the microphone and translated them into a computerized version. In this form the sounds could then be stored on a computer disk. The process of recording sounds onto a computer disk is called digitizing. This procedure was followed to digitize all of the aural examples used in the theory test. The digitized sounds stored on the disk were accessed by the theory test and played through the internal speaker of the Macintosh computer. The sound quality of the internal Macintosh speaker was adequate for the reproduction of the simple examples used in the test.

The final step in the development of the computer version of the theory test was the creation of data retrieval routines. The computer made the collection of data somewhat easier since the computer could carry out activities

demogra

test. Th

along w

choir, ar

A

student'

answer

accessib

compar

respons

on the

test adr

prepara

packag

printec

any hu

all asp

to mal

from

"Com

and g

upda

to fiv

gradu

activities in the background while the examinee continued the test. Simple demographic information was gathered during the instruction phase of the test. This information included the examinee's name and phone number, along with information concerning participation in piano lessons, band, or choir, and whether the examinee was a vocal or instrumental performer.

As the student completed each test item the computer placed the student's response in the appropriate response box on the answer sheet. The answer sheet remained hidden from the student at all times and was not accessible by the student. At the completion of the test the computer compared each response with the correct response, marked each inaccurate response, placed the number correct and incorrect into the appropriate boxes on the answer sheet, and printed the answer sheet which was collected by the test administrator. The final task performed by the computer was the preparation of all information and responses for exporting to a statistics package for analysis. To summarize, the computer administered, scored, printed results, and prepared the data for statistical analysis without requiring any human assistance other than the input provided by the student.

Repeated trials were performed on the computer test to determine that all aspects of the test presentation functioned properly. The test was checked to make sure that (a) each test item was correctly presented, (b) movement from question to question was functioning correctly including the use of the "Come Back Later" button, (c) all answer buttons were functioning properly and generating the associated response, and (d) record-keeping files were updated correctly with the demographic information and test item responses.

To assess the performance of the test in actual use, it was administered to five students of similar age and education. The five students were all graduated seniors who were actively involved in their high school music

programs
were ask
inaccura
of taking
the com
on pape
Scores f
years of
in gene

T
comple
whethe
school
which
desired

programs through choir or band. While completing the test, the students were asked to provide insights into questions they thought were unclear or inaccurately presented. They were also asked to evaluate the ease or difficulty of taking the test on the computer. All students responded that they enjoyed the computer test more than they would have enjoyed taking the same test on paper because of the accompanying animation, sound, and graphics. Scores from this pilot test ranged from 71 out of 90 for a student with thirteen years of piano experience to 29 out of 90 for a student who admitted his grades in general music or band were below average.

The test was also pilot tested with one student who had already completed first-year music theory at the college level. This was to determine whether the test content did indeed correspond to material expected of high school seniors and college freshmen. The examinee scored 80 of a possible 90 which indicated that the content parameters of the test were within the desired range.

Test Ad

T

Compu

beginni

gained

theory

require

created

search

the su

the re

admir

somet

their

music

one c

the n

the c

assig

store

CHAPTER 4 - TEST ADMINISTRATION AND RESULTS

Test Administration

The administration of the theory test took place in the Music Computer Lab at Michigan State University during registration week at the beginning of the Fall term of 1988. Due to the possibility of curricular benefits gained as a result of developing a valid predictive/diagnostic tool for music theory placement, an arrangement was made with the music theory area to require each incoming freshman to take the theory test. This arrangement created the research sample without the necessities of consent forms or a search for volunteer subjects.

Each freshman student received a School of Music newsletter during the summer of 1988 which included a brief notice informing the students of the required test along with an alphabetized schedule of times for test administration. Students were asked to report to the computer music lab sometime during the appointed hours of the day assigned to the first letter of their last name.

The theory test was installed on four Macintosh computers in the music computer lab. As the students arrived at the lab, they were assigned to one of the four machines and given a brief statement concerning the use of the mouse with the computer. All other instructions were administered by the computer. After completing the ninety test items, the students were assigned an ID number which they typed into the computer. The computer stored the test results and sent each student's item responses and final score to the pri

The test

at the en

T

First, a

test desi

by the a

Michiga

acquisi

technic

admini

was re

simply

the tes

studer

two th

beginn

appoi

incom

were

testin

subje

comp

Each

weel

to the printer where the printed copy was collected by the test administrator. The test results were removed from each computer and stored on floppy disks at the end of each test day.

Two problems were encountered during the administration of the test. First, a number of test items employed sound output from the computer. The test design called for the use of stereo headphones plugged into the computer by the administrator. However, a strike by the Clerical/Technical union of Michigan State University during the administration of the test made acquisition of the headphones impossible since the School of Music sound technician was a member of the union. This problem did not affect the administration of the test except for occasions when more than one student was responding to the aural test items. When this happened, one student simply waited until the other student finished listening.

A second problem was encountered when students did not report for the test at their assigned times. The test schedule was designed to allow the students to complete the test prior to the start of classes. Yet, approximately two thirds of the freshmen students did not report to the test center before the beginning of fall classes. Many excuses were given including forgotten appointments, non-receipt of the test notice, and failure to realize that all incoming freshmen were required to take the test. The untested students were contacted through in-class announcements and follow-up and their testing was subsequently rescheduled.

The research sample for the theory test included 59 subjects. Each subject was given an opportunity to volunteer for a retest for use in computing reliability statistics and twenty subjects were randomly selected. Each of the volunteers was retested with the same test approximately one week after the initial test.

Demogra

Al

they ans

sample

followin

a school

subjects

vocal or

demogr

Table 2

Studen

Are

Piano

Band

Choir

Majo

Demographic Characteristics of the Sample

All subjects were asked to respond to several demographic items before they answered any test items. These items were used simply as indicators of sample characteristics. The subjects were asked to answer "yes" or "no" to the following: (a) "Have you taken piano lessons?", (b) "Have you performed in a school band?", and (c) "Have you performed in a school choir?" The subjects were also asked to indicate whether their main performance area was vocal or instrumental. Table 2 shows the subjects' responses to the demographic questions.

Table 2

Student Responses to Questions of Previous Musical Experiences

	Responses	
Area of Experience	Yes	No
Piano Lessons	23	36
Band Performance	46	13
Choir Performance	33	26
	Instrumental	Vocal
Major Performance Area	42	17

Descript

In

frequen

somewh

of 85, p

standar

differen

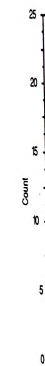
distribu

coeffici

negativ

scores

in the



Figur

Descriptive Statistics on Test Scores

Initially, the test scores of the 59 subjects were compiled into the frequency distribution seen in Table A-1 (see Appendix A). The scores were somewhat evenly distributed between the minimum of 24 and the maximum of 85, producing a range of 61. The mean for the 59 scores was 56.75 with a standard deviation of 11.97 while the median score was 55. The small difference between the mean and median indicates a comparatively even distribution. The skewness coefficient was –.074. This negative skewness coefficient indicates that a slightly greater number of student scores fell on the negative side of the mean. A more visual indication of the distribution of scores was achieved by converting the test scores to z scores and plotting them in the graph shown in Figure 1.

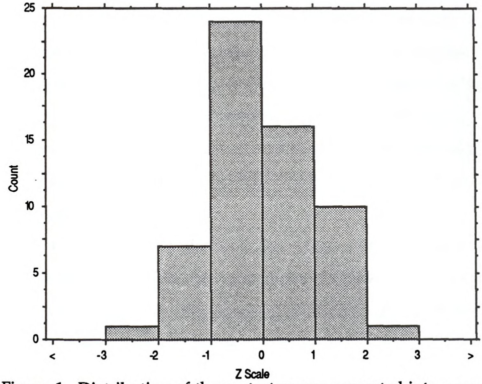


Figure 1. Distribution of theory test scores converted into z scores.

Test Iten

C

index fo

the perc

Appenc

was ma

of item

would

the sul

increas

other

of the

much

withi

of 10

sugg

midp

perc

test

the s

each

exar

corr

cod

"2".

bet

Test Item Difficulty and Discrimination Indices

Calculation of a frequency distribution analysis produced the difficulty index for each test item. The difficulty index was determined by calculating the percentage of subjects responding correctly to each item. Table B-1 (see Appendix B) displays the calculated difficulty for each item. A design decision was made in the initial stages of test development to strive for a broad range of item difficulties. Several items were included which, it was assumed, would have little or no discrimination value. In other words, almost all of the subjects were expected to respond correctly. These items were designed to increase the subjects' motivation to try to perform well on the test. On the other hand, several items were specifically included to assess the outer limits of the subjects' musical knowledge, thus the difficulty level of these was much higher. Yet, even with this goal in mind the item difficulties fell within reasonable limits. Item difficulties ranged from a low of 15% to a high of 100% and the mean difficulty was 63%. Marshall and Hales (1971) suggested that mean item difficulties should fall somewhere close to the midpoint between the percentage possible through chance and 100%. Chance percentage for the theory test was approximately 25% since the majority of the test was four-choice multiple-choice items. A mean of 63% is very close to the suggested midpoint.

Calculation of a discrimination index involved several steps. First, each subject's responses were recoded as either correct or incorrect. For example, if a test item offered four possible answers and answer "A" was the correct answer, then each subject who did not select "A" for that item was coded with a "1" while each subject who responded correctly was coded with a "2". Once this was accomplished it was possible to run a correlation analysis between each subject's test score and the coding for each test item. The purpose

item wo

I

indices.

(1971) s

discri

unde

expla

whic

for i

test

com

liste

Inte

wer

acr

purpose of the analysis was to determine whether correct answers on each item would correlate with a higher overall test score.

In general, test developers attempt to achieve large discrimination indices. However, several factors can foil this attempt. Marshall and Hales (1971) suggested the following.

> The discriminatory power of the item is influenced by a number of factors, including: the previous learning experiences of the examinees; the ability of the [question] stem (aided by the choices offered) to structure the question for the examinee; the extent of ambiguity in the item; the ability of the foils to appeal to those with incorrect or lack of knowledge; the presentation of only one best or correct answer which will appeal to the upper group; and the difficulty of the item. (p. 232)

Marshall and Hales went on to state that the greatest number of discriminations are achieved when the discrimination index is .50. This understanding of discriminations is enhanced by Nunnaly's (1961) explanation. He suggested that test creators should be suspicious of items which fall below a discrimination index of .20. There is a place in test creation for items with low discriminations as motivational tools. On the whole, a test should include items which most or all subjects can successfully complete.

The results of the discrimination calculations for the present test are listed in Table B-1. The correlations ranged from a low of -.01 to a high of .62. Interestingly, both the mean and the median for the item discriminations were calculated at .31. Table 3 shows the spread of the 90 item discriminations across various ranges.

Table 3

Numbe

were

A ne

betw

score

who

item

item

Test

feas

sele

tool

suff

the

of t

Table 3

Number of Test Item Discriminations Within Selected Ranges

Range	Number of Items
-.10 to .20	18
.21 to .40	52
.41 to .60	16
.61 to .80	2

Two test items were correctly answered by all subjects and, therefore, were not used in the analysis. One test item produced a negative correlation. A negative correlation indicates that there was an inverse correlation between that test item and final test scores. In other words, subjects who scored higher on the test were more likely to miss this item while subjects who scored lower on the test were more likely to respond correctly to this item. The discrimination index for this test item was so close to zero that the item was evaluated as non-discriminating.

Test Reliability

The test-retest method of reliability assessment was selected as the most feasible for this study. As stated earlier, twenty subjects were randomly selected to take the same test a second time. The second test administration took place approximately one week after the first administration. This gave sufficient time for the subjects to forget the details of the test while limiting the amount of time available for large changes in musical knowledge. Most of the subjects scored surprisingly close to their first test score (Table 4). A

Pearson

This is

stability

Pearson correlation analysis of test-retest scores produced a correlation of .86. This is a relatively strong correlation and seems to indicate a good level of stability between administrations of the test.

Table 4

Test-Retest Scores of 20 Students

Student #	Test	Retest
1	60	55
2	57	69
3	53	63
4	72	64
5	73	73
6	55	57
7	85	86
8	51	60
9	47	51
10	73	73
11	55	66
12	60	65
13	36	43
14	50	55
15	52	46
16	62	73
17	51	55
18	52	59
19	68	69
20	66	70

Descrip

three g

theory

two di

scores

aural

was t

deter

outsi

grad

class

subj

corr

year

wer

com

seq

for

seq

thr

dis

the

the

Descriptive Statistics on Term Grades

The criterion variables compared with the theory test scores included three grade evaluations taken from each of the three terms of the first-year theory sequence. Two of the grades were the same evaluation registered in two different grading scales.

The first grades gathered for the subjects were taken from their term lab scores. The lab portion of the theory class included drill and practice with aural skills such as sightsinging, dictation, and ear training. A separate grade was taken for this portion of the class so that a correlation analysis could determine whether the theory test was a strong predictor of success with skills outside the bounds of normal classroom assignments.

Two grades were taken which represented the subjects' final course grade. One was the final percentage based on a 100% scale as calculated by the class instructor while the second was the grade point assigned to each subject's percentage grade based on a 0.0 to 4.0 grading scale. This allowed correlation of the theory test scores with both a large and small interval scale.

All three grades were gathered for each of the three terms of the first-year music theory course. Some of the subjects who initially took the test were dropped from the study at the end of each term because they did not complete the course or elected to take subsequent terms of the theory sequence at a later date. An average grade for the entire year was calculated for each subject who completed all three terms of the first-year theory sequence. The average grades were computed by calculating the mean of the three term grades for each subject in each of the three grade criteria. Table 5 displays pertinent descriptive statistics for the three terms with divisions for the three grading areas. The average grade statistic should not be construed as the mean grade of each group of three term grades. Rather, the average grade

statisti
three t
may b
C-1, C
the ca

statistic only reflects the mean grade of the 32 subjects who completed all three terms of the music theory sequence. Hence, the average grade mean may be larger than any of the other means with which it is associated. Tables C-1, C-2, and C-3 (see Appendix C) contain the frequency distributions used for the calculation of these statistics.

Table

Descri

Theor

Grade

Lab

Fa

W

S

A

Grad

F

V

S

Gra

No

inc

aO

firs

bA

cA

Table 5

Descriptive Statistics on Grades for Three Terms of the First-Year Music Theory Sequence

Grade Area	Mean	SD	Number of Subjects
Lab			
Fall Grade	81.00	19.45	57
Winter Grade	80.42	12.73	41
Spring Grade	85.46	10.35	33
Average Grade[c]	84.96	8.61	32[a]
Grade as Percent			
Fall Grade	86.19	12.68	57
Winter Grade	86.29	8.13	41
Spring Grade	85.33	16.75	33
Average Grade[c]	87.69	7.67	32
Grade Point [b]			
Fall Grade	2.80	.98	57
Winter Grade	2.71	.91	41
Spring Grade	2.80	.93	33
Average Grade[c]	2.92	.72	32

Note. Subjects given a grade of zero in any criteria used for this table were included in these calculations but not in later correlation calculations.

[a]One subject completed the second and third terms but did not complete the first term and was not included in the average grade calculations.

[b]All grade points are calculated on a 0.00 to 4.00 scale.

[c]Averages only reflect the 32 students who completed all three terms.

0%-10
becom
appro
scale
small
variar
three

stanc
perce
lab g
grad
poss
abili
prac
do t
pers
arri
sub
be

mi
co
th
ex

The mean for all three grade criteria in Table 5 was quite high on the 0%-100% scale. The establishment of clear correlations between two criteria becomes increasingly more difficult as the mean of either criterion approaches the upper limits set for that criterion. As the mean of a given scale approaches the upper limit of that scale it indicates the presence of smaller and smaller degrees of variance between subjects and this small variance may, in turn, hinder strong correlations. The high mean of the three grade criteria may produce this effect.

Another interesting aspect of Table 5 was the way grade means and standard deviations changed with subsequent terms. The means of the grade percentages criterion were quite stable throughout the three term period. Yet, lab grade means showed marked improvement with each term while the lab grade standard deviations display significant decreases. These differences are possibly explained by the different skills needed to achieve each grade. One's ability to do normal classroom assignments is not necessarily affected by practice. Rather, the individual either chooses to do the assignment or not to do the assignment. A lab class, however, emphasizes listening and performance skills which may be quite undeveloped when the student arrives as a freshman. Over the course of the year, many students are able to substantially improve their skills in these areas and this improvement would be reflected in increased grade means.

The changes in the mean and standard deviation of the lab grades might also be the result of the failure of weaker students to continue in the course. The withdrawal of weaker students results in a higher mean since those students' low scores are no longer included in the calculation. The exclusion of these low scores is also reflected in a smaller standard deviation since t

is decr

Corre

The g

predic

was c

each

test v

grad

comp

pres

subj

stud

skill

norr

acti

stin

wit

the

dis

since the size of the variance between the high and low extremes of the group is decreased.

Correlation of the Theory Test and Class Grades

The goal of this research study was to design a tool which was useful as a predictive/diagnostic test of success in first-year theory. A correlation analysis was completed between the theory test and the course grades gathered from each term to evaluate the success of the test in fulfilling the stated goal. The test was compared individually with each lab grade, percentage grade, and final grade for each term. Then the test was compared with a three term average grade in each of the three areas. Table 6 presents the final correlation statistics.

The lowest correlations occurred between the theory test and the subjects' lab grades. This is perhaps an expected result. In a theory lab, students are drilled in ear training, sightsinging, dictation, and other aural skills. Only one tenth of the theory test was devoted to any of the activities normally included in a theory lab. All the questions incorporating lab-type activities were specifically focused upon aural recognition of a musical stimulus. The student demonstrated recognition by matching the stimulus with a correct response on the screen. In general, a low correlation between the theory test and the subjects' lab scores was expected due to the dissimilarity of content areas.

Table

Correl

Grade

Grade
Lab C
Perc
Fin

No
thi
a A

Table 6

Correlational Statistics for Comparison of the Theory Test With Classroom Grades

Grade Area	Subjects	Correlation
Lab Grade		
Fall	56	.29
Winter	41	.25
Spring	33	.24
Average[a]	32	.32
Percentage Grade		
Fall	57	.41
Winter	41	.43
Spring	33	.36
Average[a]	32	.40
Final Grade Point		
Fall	54	.42
Winter	41	.47
Spring	33	.43
Average[a]	32	.51

Note. Subjects who were given a grade of zero in any of the criteria used for this table were not included in correlation calculations.

[a]Averages only reflect the 32 students who completed all three terms.

shoul
correl
correl
comp
grad
it is a

corre
decli
spri
less
resu
stro
stre

bet
4.0
poi
var
str
ins
rec
gr

st
ac

As noted in the discussion of Table 5, the average grade correlations should not be construed as the mean correlation of each group of three term correlations. Rather, the average grade correlation only reflects the correlation of the theory test with the average grades of the 32 subjects who completed all three terms of the music theory sequence. Hence, the average grade correlation may be larger than any of the other correlations with which it is associated.

Both the percentage grade correlation and the final grade point correlation displayed relative stability. An interesting observation was the decline in correlation strength of the percentage grade calculated for the spring term. One possible hypothesis for this decline was that students were less motivated during the spring months and sagging grades were a natural result of slackening motivation. This grade "slump" appeared to affect the stronger students more than the weaker students thereby lessening the strength of the correlation.

Perhaps the most important aspect of Table 6 was the correlation between the theory test and the average of final grade points based on a 0.0 to 4.0 scale. The strength of the correlation (r=.51) was surprising since the grade point scale was the least sensitive of the interval scales used as criterion variables. However, it was this correlation which carried the greatest strength. Both the lab grade and percentage grade were calculated by the instructor but only the final grade point was recorded on the students' records. Ideally, then, the theory test should correlate most strongly with the grade which will eventually show up on the students' transcripts.

One drawback repeatedly presented itself during interpretation of the statistical data. The distribution of the theory test scores was somewhat even across a large range while the distribution of grades in all three grade criteria was g

varia

theor

corre

Fall
(Pe

Fig
nur

wa

no

of

de

va

sk

was greatly skewed toward the high end of the scale. I believe that a larger variance in lab grades would produce a stronger correlation between the theory test and student grades. Figure 2 demonstrates this problem with a correlation between theory test scores and the fall lab grades.

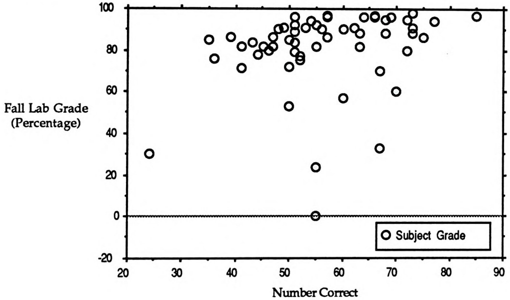


Figure 2. Correlation scattergram between Fall term lab grade percentages and number correct on the theory test.

At this juncture the three-term averages of subject final grade points was once again noteworthy since the scattergrams for these correlations did not indicate such a high-end skewness (Figure 3). Perhaps the smaller range of the grade point scale made it less sensitive to this problem. Figure 3 also demonstrates that the heavier weight of classroom grades and the grade variance achieved by the classroom instructor may offset the high-end skewness of lab grades.

Ave
Gr

Figu
corr

jud

As

nee

tes

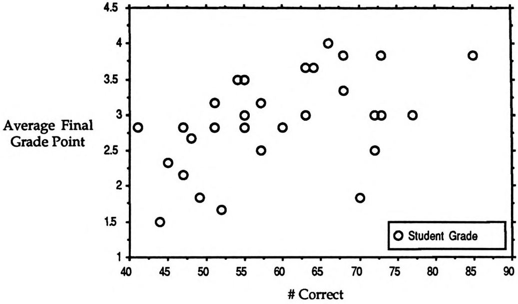


Figure 3. Correlation scattergram between average final grades and number correct on the theory test.

Overall, the test score distribution and the correlation statistics were judged successful for the first iteration of a new measurement instrument. As expected, some of the test items require some adjustment while others need to be replaced. However, no major design flaws were discovered in the test.

Probl

of th
techn
such
did

secc
to r
of t
A r
Ho
sim
firs
rea
hav
scl
mu
an
m
cc
C

CHAPTER 5 - DISCUSSION

Problems With The Study

Several problems which arose during the design and implementation of the study were mentioned in previous chapters. Some, like the technicians' strike at the university, were unavoidable but minor. Others, such as the failure of subjects to report at assigned times, were irritating but did not adversely affect the study.

One minor flaw in the research design became apparent after the second term of course grades was gathered. The problem involved a failure to recognize that freshmen students were not required to take all three terms of the first-year music theory sequence during their first year of enrollment. A number of subjects dropped out of the theory sequence after each term. However, it was inaccurate to label a student as unsuccessful in music theory simply on the basis of that student's failure to complete all three terms of the first-year theory sequence. It was not possible to ascertain each student's reason for delaying completion of the sequence, and these reasons might have included failure to do well in the class, financial considerations, or scheduling difficulties. Any students who did not complete a given term of music theory or did not enroll were simply deleted from the statistical analysis for that term. Students who did not complete all three terms of music theory were included in the statistical analysis for terms which they did complete and this accounts for the diminishing sample size noted in Chapter 4.

with
make
use v
the s
size
subje
subj
fresh
subj
that
wou
Oth

ins
fre
pro
bet
oth
to
rev
Co

ite
In
no
n

The loss of subjects from successive terms produced another problem with the study. The original sample size of 59 subjects was large enough to make some specific generalizations concerning the music theory test and its use with freshmen music students at Michigan State University. However, the strength of the generalizations was somewhat weakened as the sample size decreased. It was not possible to remedy this situation since all of the subject pool was used up for the 1988-1989 academic year and the available subject population is only replenished once per academic year when new freshmen students arrive on campus. The problem of limited number of subjects was addressed by my Doctoral Guidance Committee which concluded that the project should continue but that follow up work with more subjects would be a profitable endeavor at a later time.

Other Applications

One important consideration of this test is its usefulness to other institutions. The validity of this study is only claimed when it is used with freshmen music majors at Michigan State University. Applicability to theory programs in other institutions, however, may be based on the similarities between the traditional theory program at Michigan State University and other theory programs. It is premature at this point in the test development to try to generalize the results of this study to other theory programs. Further revision of the test may allow this in the future.

Conclusions

The development of useful testing tools requires a number of iterations and the theory test under discussion in this paper was no exception. In general, the test performed as expected. Several meaningful findings were noted. First, the correlation between the test and the subjects' lab grade was not strong enough to warrant use of the test as a predictor of success in this

portio
the st

test i
unco
need
that
subje
disc

theo
thar
imp
topi
stro

the
tes
str
pa
pr

pr
ev
st
th
cr

portion of the first-year music theory sequence. Suggestions for enhancing the strength of this correlation are included in the succeeding section.

The test items themselves performed reasonably well. The range of test item difficulties was well distributed although some weaknesses were uncovered during the analysis of item discrimination. A number of items need reevaluation and some items should be discarded, although I believe that some items with low discrimination ability are necessary to increase subject motivation. I was pleased to note the small number of negatively discriminating test items.

Perhaps the most striking result of the study was the discovery that the theory test correlated more strongly with the less sensitive grade point scale than it did with the percentage grade scale. As stated previously, this is an important finding since the grade point scale is used for student records. A topic for continued study is whether the grade point scale continues as the strongest criterion as the test improves through successive revisions.

In general, the theory test appeared to be an encouraging start toward the development of a useful diagnostic/predictive tool for music theory. The test items, in most instances, functioned successfully. Evaluation of the strength of the reliability statistic and the acknowledgement by the theory panel of experts that the test content validity was sufficiently high indicate the presence of a strong measurement foundation for future revisions of the test .

The computer implementation of the test was very successful. No problems arose with either the software or the hardware. This study did not evaluate student responses to computerized testing but it appeared that the students enjoyed working with the computers. This was especially evident in the number of questions asked concerning the development process used in creating the test and placing it on the computers.

adeq

to wa

was

large

for a

head

Sug

test

cap

whi

sele

abi

stre

Ho

stre

sin

co

ob

co

in

es

fr

th

th

The number of computers available for test administration was adequate for the number of students taking the test and students did not have to wait for a computer to become available. However, the test administration was spread across an entire week. More computers would be necessary if larger numbers of students were required to take the test or if the time given for administration of the test was shortened significantly. Personal headphones for each student would also be essential.

Suggestions for Further Research and Improvement

A number of avenues are available for continued development of this test to strengthen its current usefulness and enhance its assessment capabilities. Of primary import is the refinement of current test items. Items which are used as non-discriminating, motivational questions should be selected and retained. All other items which have low discriminating abilities must be rewritten. Some questions might be salvaged by strengthening the incorrect choices, thereby increasing the item difficulty. However, the item difficulty rating for some weak items is already sufficiently strong. These items might have to be replaced by new questions which cover similar material.

The test might be enhanced by the addition of some additional music content areas. Several areas suggested during the development of test objectives included form, syncopation, and twentieth century theoretical concepts such as bitonality and tone clusters. These areas were not included in the initial test form because they fell outside the content parameters established by the first-year theory instructor. Perhaps students would benefit from the inclusion of some or all of these content areas in the first-year theory curriculum. If new content areas were added to the curriculum, then those new areas could be used to generate new questions for the theory test.

Test

adde

minu

willi

test

wou

num

sign

add

lab

mig

stu

dic

sca

rep

qu

pi

no

th

an

mi

re

du

pr

te

Test length must be carefully considered before any new content areas are added. Completion of the test in its current form requires approximately 50 minutes. Increasing the length of the test is likely to decrease a student's willingness and motivation to do his or her best because the duration of the test stretches the limits of the student's attention span. Therefore, new items would have to replace already constructed items. This, in turn, decreases the number of test items devoted to other content areas which are perhaps more significant in the overall theory curriculum. The trade-offs involved with adding the new test items might be prohibitive.

The statistics in Chapter 4 demonstrated the low correlation between lab grades and scores on the theory test. It is possible that this correlation might be strengthened through the addition of more items which assess the student's performance skills in areas such as sightreading, ear training, and dictation. In the current version of the test, the student is asked to identify scales, chords, and intervals played by the computer. This may be an adequate representation of listening items. Examples of other test areas might include questions in which the student is asked to press the keys of an on-screen piano keyboard to "play" a melodic line or the student might be asked to drag notes onto a staff to notate a melodic line played by the computer. Both of these examples are easily done with the computer but involve a much greater amount of time for student responses. Again, the problem of test duration might prohibit expansion into some areas which, by their presence, could result in a stronger measurement tool.

A final suggestion for enhancing the test is not affected by the test duration problems listed above since it replaces one method of item presentation with another. The digitized sounds used to present the aural test items were quite adequate. However, the use of MIDI connections to a

synth
of the
proto
instr
conn
poss
in th
inco
enha
harc
time
MII
mu
incr
mig
not
kno
per
anc
eq

Ea
be
ac
or

su

synthesizer keyboard would greatly increase the quality of the musical output of the test. MIDI or Musical Instrument Digital Interface is an established protocol for transferring musical data from one instrument to another. The instruments can be synthesizers, computers, or even guitars with appropriate connections. Connecting a computer to a synthesizer via MIDI makes it possible to play the synthesizer from the computer. The aural examples used in the theory test would be of a much higher quality if MIDI technology was incorporated to produce the examples on a synthesizer. The drawback of enhancing the theory test with MIDI technology lies in the increased hardware and software demands. On the software side there is a much longer time involvement dedicated to programming the theory test to produce the MIDI signals which the synthesizer requires. On the hardware side there is a much greater investment in equipment required to administer the test. The increased amount of equipment also increases the number of problems which might arise during test administration. For example, the synthesizer might not be properly set to receive MIDI information and most students would not know how to remedy this problem. Problems such as this can lead to lower performance on the test. In its present form the theory test is self contained and does not require the test administrator to check any peripheral equipment.

Each enhancement listed above can, in some way, make the test better. Each enhancement also has associated drawbacks which must be considered before implementation. Only further development of the theory test and administration to new samples will adequately demonstrate the value of any or all of the possibilities.

In summary, I have created a test for diagnosing and predicting the success of first-year music theory students at Michigan State University. The test w

subjec

satisfa

probl

curric

Univ

follov

test was implemented on the Macintosh computer and administered to 59 subjects without incident. The statistical results calculated were quite satisfactory for the first iteration of a test instrument with the exception of the problems noted in this chapter. Other institutions whose first-year theory curriculum is somewhat parallel with the curriculum at Michigan State University might be interested in the results of this research project and any follow-up research completed at a later time.

Table

Frequ

APPENDIX A - FREQUENCY DISTRIBUTION OF TEST SCORES

Table A-1

Frequency Distribution for 59 Theory Test Scores

Score	Frequency
85	1
77	1
75	1
73	4
72	2
70	1
69	1
68	2
67	2
66	2
64	1
63	2
62	1
60	2
59	1
57	3
56	1
55	5
54	1
53	1
52	2
51	5
50	3
49	1
48	1
47	2
46	1
45	1
44	1
43	1
41	2

Table

Table A-1 (cont'd.).

39	1
36	1
35	1
24	1

APPENDIX B - ITEM DIFFICULTIES AND DISCRIMINATIONS

Table B-1

Test Item Difficulties and Discriminations Displayed as Percentages

	Indices	
Question #	Difficulty	Discrimination
1	.98	.13
2	.83	.31
3	.83	.32
4	.90	.27
5	1.00	—
6	.83	.38
7	.68	.29
8	.64	.33
9	.93	.16
10	1.00	—
11	.70	.10
12	.88	.20
13	.80	.35
14	.90	.36
15	.56	.42
16	.54	.29
17	.51	.35
18	.70	.07
19	.86	.38
20	.70	.14
21	.39	.54
22	.17	.26
23	.46	.31
24	.78	.30
25	.64	.41
26	.75	.34

Table B-1 (cont'd.).

27	.85	.50
28	.85	.34
29	.95	.14
30	.48	.06
31	.85	.37
32	.70	.37
33	.44	.41
34	.85	.24
35	.90	.39
36	.48	.30
37	.25	.29
38	.59	.37
39	.46	.25
40	.27	.26
41	.41	.21
42	.59	.37
43	.20	.28
44	.29	.45
45	.54	.34
46	.20	-.01
47	.54	.28
48	.85	.30
49	.64	.13
50	.44	.24
51	.80	.50
52	.61	.17
53	.37	.18
54	.83	.42
55	.88	.31
56	.85	.14
57	.64	.43
58	.73	.16
59	.42	.41
60	.46	.43
61	.68	.31
62	.71	.31
63	.80	.18
64	.70	.50
65	.51	.62
66	.54	.44
67	.42	.43
68	.93	.53
69	.85	.32
70	.76	.37
71	.90	.29

Table B-1 (cont'd.).

72	.70	.25
73	.49	.26
74	.15	.20
75	.75	.27
76	.22	.25
77	.73	.32
78	.27	.34
79	.64	.60
80	.70	.43
81	.31	.36
82	.49	.55
83	.53	.29
84	.75	.40
85	.44	.38
86	.53	.22
87	.58	.26
88	.24	.17
89	.39	.40
90	.92	.10

APPENDIX C - FREQUENCY DISTRIBUTIONS OF GRADES

Table C-1

Frequency Distribution for Student Lab Grades in Music Theory

	# of Subjects			
Grade Percentage	Fall	Winter	Spring	Average
99	0	0	1	0
98	1	2	1	0
97	3	0	0	2
96	5	0	3	1
95	2	1	2	0
94	2	1	2	0
93	0	1	1	1
92	3	2	1	2
91	4	0	0	1
90	3	1	4	5
89	1	0	1	1
88	3	5	3	0
87	0	3	1	0
86	4	0	1	2
85	2	3	1	3
84	2	2	0	1
83	0	1	0	2
82	5	3	0	2
81	0	0	0	1
80	2	2	2	1
79	1	1	1	1
78	1	0	0	1
77	1	1	0	1
76	1	3	0	0
75	1	0	1	0
74	0	0	2	0

Table C-1 (cont'd.).

73	0	1	0	0
72	1	0	1	0
71	1	1	1	0
70	1	1	2	1
67	0	1	0	1
63	0	1	0	0
60	1	0	0	0
59	0	1	0	0
58	0	0	0	1
57	1	0	0	0
56	0	0	1	0
55	0	1	0	1
53	1	0	0	0
47	0	1	0	0
43	0	1	0	0
33	1	0	0	0
30	1	0	0	0
24	1	0	0	0
0	1	0	0	0

Table C-2

Frequency Distribution for Student Percentage Grades in Music Theory

	# of Subjects			
Grade Percentage	Fall	Winter	Spring	Average
99	1	0	0	0
98	2	0	1	0
97	4	2	2	2
96	0	2	3	2
95	4	1	1	1
94	5	2	0	3
93	3	2	1	0
92	3	4	3	1
91	5	3	5	2
90	1	1	0	3
89	1	3	0	4
88	6	3	2	2
87	2	2	2	2
86	0	1	2	1
85	0	3	1	1
84	2	0	2	0
83	4	1	0	0
82	3	0	0	2
81	2	2	0	0
80	1	1	1	0
79	0	1	0	0
78	1	0	3	0
77	0	0	0	2
76	1	0	3	2
75	1	1	0	0
74	2	2	0	0
72	0	1	0	0
70	0	1	0	0
69	0	1	0	0
68	0	1	0	0

Table C-2 (cont'd.).

63	0	0	0	1
58	0	0	0	1
54	1	0	0	0
37	1	0	0	0
33	1	0	0	0
0	0	0	1	0

Table C-3

Frequency Distribution for Student Final Grades in Music Theory

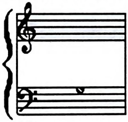
	# of Subjects			
Grade Point	Fall	Winter	Spring	Average[a]
4.0	7	4	6	2
3.5	15	9	5	7
3.0	13	10	7	8
2.5	8	7	7	8
2.0	7	4	4	2
1.5	4	2	3	5
1.0	0	5	0	0
0.5	0	0	0	0
0.0	3	0	1	0

[a]The numbers in the Average column include all subjects whose final grade is less than the grade indicated in the row above the selected row and greater than or equal to the grade label for the selected row.

APPENDIX D - COLMAN THEORY TEST

©1990 by James P. Colman. All Rights Reserved.

1. The note shown below is written in which clef?

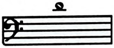


a. Treble
b. Bass

2. When four-part harmony is written on the Grand Staff the alto part is usually written in which clef?

a. Treble
b. Bass

3. Which of the answers below correctly identifies this note?

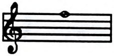


a. A
b. B
c. C
d. D

4. Which of the answers below correctly identifies this note?

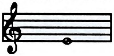
a. E-flat
b. F-sharp
c. G-sharp
d. A-flat

5. Which of the answers below correctly identifies this note?



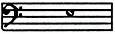
a. F
b. G
c. A
d. B

6. If this note were lowered one octave and written in the bass clef which of the choices below demonstrates the correct notation?



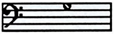
a. 

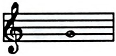
b. 

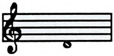
c. 

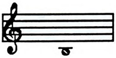
d. 

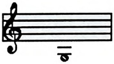
7. If this note were rewritten in the treble clef at the same pitch level which of the choices below demonstrates the correct notation?



a. 

b. 

c. 

d. 

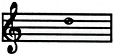
8. How many half-steps are contained in the interval from C# to F?

a. 3
b. 4
c. 5
d. 6

9. When an individual plays or sings the notes E and F are they performing a half-step or a whole step?

a. Half step
b. Whole step

10. Which of the accidental signs below will raise this note a chromatic half-step?



a. ♮ b. ♯

c. ♭ d. ♭♭

11. To lower this note a chromatic half-step, which of the accidental signs below must replace the current sign?

a. ♮ b. ♯

c. ♭ d. ♭♭

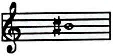
12. Which of the accidental signs below will raise this note a chromatic half-step?

a. ♮ b. ♯

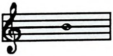
c. ♭ d. ♭♭

13. Of the choices given below, which one is another way of notating this pitch?

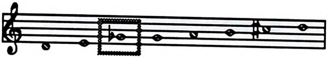
a. 

b. 

c. 

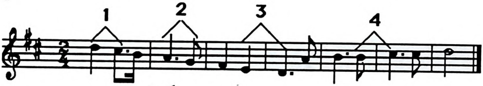
d. 

14. The following scale contains all the correct pitches. The indicated pitch is notated incorrectly. Which of the choices below demonstrates the correct notation?



a. 
b. 
c. 
d. 

15. In the musical example below, which numbered interval is a diatonic half-step?

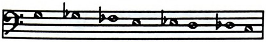


a. 1
b. 2
c. 3
d. 4

16. In the musical example below which numbered interval is a chromatic half-step?

a. 1
b. 2
c. 3
d. 4

17. Which one of the choices below contains a portion of a correctly notated chromatic scale?

a. 
b. 
c. 
d. 

18.
the

a

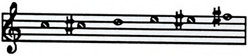
19.
acc

20.
giv

21.

18. Which one of the following selections contains the correct conclusion of the chromatic scale started below?

a.  b. 

c.  d. 

19. Only one of the notes shown below correctly completes this measure according to the meter signature. Which one?

a.  c. 

b.  d. 

20. Which of the meter signatures shown below is the best match with the given musical example?

a. $\frac{4}{4}$ c. $\frac{3}{4}$

b. $\frac{9}{8}$ d. $\frac{6}{8}$

21. Select the answer which contains the best example of compound meter.

a.  b. 

c.  d. 

22.
note

23.
min

24.
dur

25.

26.
dur

27.

22. A half-note triplet is equivalent to which of the following unsyncopated note groups?

a. ♩ ♩ ♩ b. 𝅗𝅥 𝅗𝅥

c. 𝅗𝅥 𝅗𝅥 𝅗𝅥 d. 𝅝 𝅗𝅥

23. If a metronome marking indicates the rate of a half-note at 120 beats per minute, what is the rate of a whole-note.

a. 60
b. 120
c. 180
d. 240

24. Which of the following choices demonstrates the correct sequence of note durations from longest to shortest?

a. 𝅗𝅥 ♩ 𝅜 ♪ b. 𝅜 ♩ ♪ 𝅘𝅥𝅯

c. 𝅝 ♩ 𝅗𝅥 ♪ d. 𝅝 𝅜 𝅗𝅥 ♪

25. How many thirty-second notes does it take to equal a dotted quarter-note?

a. 6
b. 8
c. 10
d. 12

26. Which of the following choices demonstrates the correct sequence of rest durations from longest to shortest?

a. 𝄻 𝄼 𝄽 𝄾 b. 𝄼 𝄻 𝄾 𝄿

c. 𝄻 𝄼 𝄿 𝄾 d. 𝄼 𝄽 𝄿 𝄾

27. Which rest given below correctly completes this measure?

a. 𝄼 b. 𝄾

c. 𝄽 d. 𝄻

28. C
note

a

29.
one'

a

30.
tec

31.
be

28. One of the choices below does not conform to the basic rules for adding note stems. Which one?

a. 

b. 

c. 

d. 

29. One of the choices below demonstrates incorrect use of note flags. Which one?

a. 

b. 

c. 

d. 

30. One of the choices below demonstrates incorrect note beaming techniques. Which one?

a. 

b. 

c. 

d. 

31. How many dotted-quarter notes are required to fill in the empty measure below?

a. 2
b. 3
c. 4
d. 5

32. C
of too

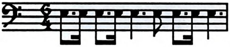
33.
belo

34.
tiec

35

32. One of the measures below is rhythmically incorrect due to the inclusion of too many dotted rhythms. Select the incorrect measure.

a.  c. 

b.  d. 

33. Based on the indicated meter signature, which of the dotted-rests shown below would add the number of beats necessary to complete this measure.

a.  b. 

c.  d. 

34. Select the single note which has a duration equal to that signified by the tied notes shown below if written in the same meter.

a.  c. 

b.  d. 

35. Which group of tied notes below has the longest duration?

a.  c. 

b.  d. 

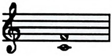
36.
inte
clic

37.
int
cli

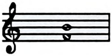
38
in
cl

3

36. Aural Question: Click the button marked "Play It". After hearing the interval select the answer which contains the interval you heard. (You may click the button three times.)

a. 
b. 
c. 
d. 

37. Aural Question: Click the button marked "Play It". After hearing the interval select the answer which contains the interval you heard. (You may click the button three times.)

a. 
b. 
c. 
d. 

38. Aural Question: Click the button marked "Play It". After hearing the interval select the answer which contains the interval you heard. (You may click the button three times.)

a. 
b. 
c. 
d. 

39. Select the name which identifies the interval shown below?

a. Major 3rd
b. Minor 3rd
c. Major 4th
d. Perfect 4th

40.
inter
acco

41.
int

42

4
m

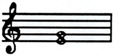
40. It is possible to form the pictured interval by combining two smaller intervals. Which answer names two small intervals which would accomplish this task?

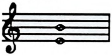
a. 3rd and 5th
b. 4th and 7th
c. 5th and 7th
d. 4th and 5th

41. Is the interval shown below a major, minor, augmented, or diminished interval?

a. Major
b. Minor
c. Diminished
d. Augmented

42. Select a perfect interval from the answers below.

a. 

c. 

b. 

d. 

43. In this musical example, is the quality of the marked melodic interval major, minor, augmented, or diminished?

a. Major
b. Minor
c. Diminished
d. Augmented

44. I
majo

45.
oct

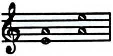
46
d

4

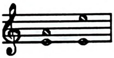
44. In this musical example is the quality of the marked melodic interval major, minor, perfect, or augmented?

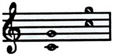
a. Major
b. Minor
c. Diminished
d. Augmented

45. Which of the answers below demonstrates interval inversion at the octave?

a. 

c. 

b. 

d. 

46. Is the inversion of this interval major, minor, augmented, or diminished?

a. Major
b. Minor
c. Diminished
d. Augmented

47. Select the correct inversion for a minor third.
a. Major 5th
b. Major 6th
c. Minor 6th
d. Minor 7th

48. Select the name of the major scale which has F as its fifth degree.
a. A
b. B
c. B-flat
d. C-flat

49. S
degre

50.

51.

52
se
be

5

49. Select the name of the minor scale which has C-sharp as its seventh degree.

a. D
b. D-flat
c. E
d. E-flat

50. Choose the name of the scale which has both E-flat and F-sharp in it.

a. c minor
b. d minor
c. e-flat minor
d. g minor

51. Which of the scales shown below is not a major scale?

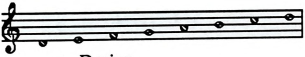
a. 

c. 

b. 

d. 

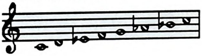
52. The terms "major" and "minor" indicate a musical mode. There are several other musical modes. Which mode is indicated by the example below?

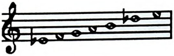


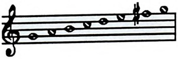
a. Dorian
b. Phrygian
c. Lydian
d. Mixolydian

53. Which of the scales shown below is not a diatonic scale.

a. 

b. 

c. 

d. 

54. V
scale'

55.

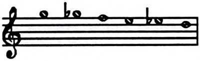
56
W
H

5
s

54. Which of the scales below is not a correctly notated portion of a chromatic scale?

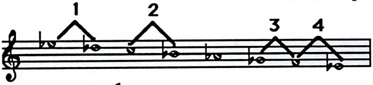
a. 

b. 

c. 

d. 

55. One of the numbered labels on this scale marks a half step. Which one?

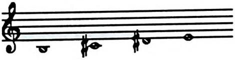


a. 1
b. 2
c. 3
d. 4

56. A major scale contains an ordered sequence of half-steps and whole-steps. Which answer below contains the proper sequence? (W = Whole step/H = Half step)

a. WWWHWWW
b. WWHWWWH
c. HWWHWWW
d. WWHWHWH

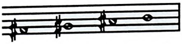
57. Select the answer which contains the second half of the major scale started below.



a. 

c. 

b. 

d. 

58. T

59.
min

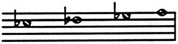
60.
ha

6
n

58. The first half-step in a minor scale occurs between which scale degrees?

a. 1 and 2
b. 2 and 3
c. 3 and 4
d. 4 and 5

59. Select the answer which contains the second half of the ascending natural minor scale started below.

a. 
b. 
c. 
d. 

60. Select the answer which contains the second half of the ascending harmonic minor scale started below.

a. 
b. 
c. 
d. 

61. Select the answer which contains the second half of the ascending melodic minor scale started below.

a. 
c. 
b. 
d. 

62.
asce

63.
scal

64.
to t

65.
to

66
to

6
t

62. Which form of the minor scale does not use the same notes for its ascending and descending portions?

a. Natural
b. Melodic
c. Harmonic

63. Which form of the minor scale uses exactly the same pitches as the major scale with the same key signature?

a. Natural
b. Melodic
c. Harmonic

64. Each degree of a scale is given a descriptive label. What is the label given to the first note of a scale?

a. Tonic
b. Dominant
c. Mediant
d. Subdominant

65. Each degree of a scale is given a descriptive label. What is the label given to the fifth note of a scale?

a. Supertonic
b. Subtonic
c. Subdominant
d. Dominant

66. Each degree of a scale is given a descriptive label. What is the label given to the major seventh scale degree?

a. Submediant
b. Dominant
c. Supertonic
d. Leading Tone

67. Each degree of a scale is given a descriptive label. What is the label given to the fourth note of a scale?

a. Tonic
b. Dominant
c. Subdominant
d. Mediant

68. Aural Qu
scale select th
listen three t

69. Aural Q
scale select t
listen three

70. Aural Q
scale select
listen three

71. Aural
scale selec
listen thre

72. Selec
with fou

73. Wha

68. Aural Question: Click the button marked "Play It". After listening to the scale select the answer which contains the type of scale played. You may listen three times.

a. Major
b. Minor
c. Chromatic

69. Aural Question: Click the button marked "Play It". After listening to the scale select the answer which contains the type of scale played. You may listen three times.

a. Major
b. Minor
c. Chromatic

70. Aural Question: Click the button marked "Play It". After listening to the scale select the answer which contains the type of scale played. You may listen three times.

a. Natural Minor
b. Harmonic Minor
c. Melodic Minor

71. Aural Question: Click the button marked "Play It". After listening to the scale select the answer which contains the type of scale played. You may listen three times.

a. Major
b. Harmonic Minor
c. Melodic Minor
d. Chromatic

72. Select the answer which contains the name of the major and minor keys with four sharps.

a. E major, C# minor
b. E major, G minor
c. B major, C# minor
d. B major, G minor

73. What is the key signature for the key of E-flat minor?

a. 

c. 

b. 

d. 

74. Select the

75. Select th

76. Select t

77. Select

78. Select
musical ex

79. If the
the two

74. Select the parallel minor for the key of G major.

a. E minor
b. B minor
c. G minor
d. A minor

75. Select the relative minor for the key of A-flat major.

a. C minor
b. B-flat minor
c. G-flat minor
d. F minor

76. Select the parallel major for the key of B minor.

a. G major
b. A major
c. B major
d. D major

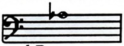
77. Select the relative major for the key of D minor.

a. B-flat major
b. C major
c. E-flat major
d. F major

78. Select the answer which identifies the most likely choice of key for the musical excerpt below.

a. D minor
b. F major
c. B-flat major
d. C major

79. If the note shown below is the third of the chord, which answer contains the two notes necessary to form a minor triad?



a. E and B
b. F-flat and C-flat
c. F and C
d. F-sharp and C-sharp

80. If this nc
contains an

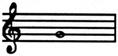
81. What t

82. The ke

83. It is p

84. Whi

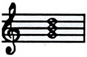
80. If this note was assigned the root position of a chord, which answer contains an augmented triad built on this note?



a. 

b. 

c. 

d. 

81. What two small intervals, when combined, form a diminished triad?

a. Major 3rd, Major 3rd
b. Major 3rd, minor 3rd
c. minor 3rd, minor 3rd
d. minor 3rd, perfect 3rd

82. The key of D-flat major does not contain which of the following triads?

a. 

b. 

c. 

d. 

83. It is possible to construct this chord in which major keys?

a. E, B, D
b. E, B, F-sharp
c. B, A, F-sharp
d. E, B, A

84. Which major key has this triad built on the third scale degree?

a. E-flat
b. D-flat
c. C-flat
d. B-flat

85. What

86. By tra
change th

87. Aural
the examp
hear. Yo

88. Aural
the examp
hear. Yo

89. Aura
the exam
hear. Yo

85. What is the bottom note of the first inversion of this chord?

a. G-flat
b. B-flat
c. D-flat

86. By transposing the top note of this chord down one octave, would you change the chord to root position, first inversion, or second inversion?

a. Root position
b. First inversion
c. Second inversion

87. Aural Question: Click on the button marked "Play It". After listening to the example select the answer which identifies the quality of the chord you hear. You may listen three times.

a. Major
b. Minor
c. Diminished
d. Augmented

88. Aural Question: Click on the button marked "Play It". After listening to the example select the answer which identifies the quality of the chord you hear. You may listen three times.

a. Major
b. Minor
c. Diminished
d. Augmented

89. Aural Question: Click on the button marked "Play It". After listening to the example select the answer which identifies the quality of the chord you hear. You may listen three times.

a. Major
b. Minor
c. Diminished
d. Augmented

90. Aural Q
the example
hear. You n

90. Aural Question: Click on the button marked "Play It". After listening to the example select the answer which identifies the quality of the chord you hear. You may listen three times.

a. Major
b. Minor
c. Diminished
d. Augmented

Aliferis, J.

Level.

Aliferis, J.

in mus

Arenson,

grades

Resea

Arnold, C

1-2).

Atkinson

Ball, C. I

of a c

Bejar, I.

Meas

Berk, R.

Balti

Bloom,

cour

Glen

Boyle, J

expe

BIBLIOGRAPHY

Aliferis, J. (1947). The Aliferis Music Achievement Test - College Entrance Level. St. Paul, MN: University of Minnesota Press.

Aliferis, J. & Stecklein, J. E. (1953). The development of a college entrance test in music achievement. Journal of Research in Music Education, 1, 83-96.

Arenson, M. A. (1983). The validity of certain entrance tests as predictors of grades in music theory and ear training. Bulletin of the Council for Research in Music Education, 75, 33-39.

Arnold, C. (1984). Organ Literature: A Comprehensive Survey (2nd ed., vols. 1-2). Metuchen, NJ: The Scarecrow Press.

Atkinson, B. (1988). HyperCard. Cupertino, CA: Apple Computer.

Ball, C. H. (1964). The application of an empirical method to the construction of a college entrance test in music. Dissertation Abstracts, 26, 404.

Bejar, I. (1984). Education diagnostic assessment. Journal of Educational Measurement, 21, 175-189.

Berk, R. A. (1980). Criterion-referenced measurement: the state of the art. Baltimore: Johns Hopkins University Press.

Bloom, B. S. & Peters, F. R. (1961). The use of academic prediction scales for counseling and selecting college entrants. New York: The Free Press of Glencoe.

Boyle, J. D., & Radocy, R. E. (1987). Measurement and evaluation of musical experiences. New York: Schirmer Books.

Brand, M. &
of error-
Caimi, F. J.
motivat
estimati
for Rese
Kruege
Chadwick,
at the
Resear
Capps, S.
Comp
Chevallar
relatec
Colwell,
Engle
Eberly, C
asses
Cour
Ernest, I
majo
Fletcher
tests
Mat
Gay, J.

Brand, M. & Burnsed, V. (1981). Music abilities and experiences as predictors of error-detection skill. Journal of Research in Music Education, 29, 91-96.

Caimi, F. J. (1984). An investigation of the relationships between certain motivational variables and selected criterion measures useful for estimating success at high school band directing. Bulletin of the Council for Research in Music Education, 77, 52-59. (Reviewed by Reynold Krueger).

Chadwick, C. S. (1976). The prediction of success in student teaching in music at the University of Southern California. Bulletin of the Council for Research in Music Education, 49, 49-54. (Reviewed by Reynold Krueger).

Capps, S. (1988). MacRecorder soundedit. Berkeley, CA: Farillon Computing.

Chevallard, P. C. (1982). Intonational performance as predicted by subject-related variables. Dissertation Abstracts, 43, 3747-8A.

Colwell, R. (1970). The evaluation of music teaching and learning. Englewood Cliffs, NJ: Prentice-Hall.

Eberly, C. G. & Cech, E. J. (1986). Integrating computer-assisted testing and assessment into the counseling process. Measurement and Evaluation in Counseling and Development, 19, 18-26.

Ernest, D. J. (1970). The prediction of academic success of college music majors. Journal of Research in Music Education, 18, 273-276.

Fletcher, P. & Collins, M. A. J. (1986). Computer-administered versus written tests - advantages and disadvantages. Journal of Computers in Mathematics and Science Teaching, 6, 38-43.

Gay, J. (1989). SuperPaint 2.0. San Diego, CA: Silicon Beach Software.

Gordon, E.

aptitude

Educatio

Gordon, E.

with col

in Musi

Gordon, E.

aptitud

for Res

Gordon, E

interm

Resea

Gordon,

study

Profi

Gregory

Jour

Gronlu

Nev

Gronlu

Ne

Harde

an

Hard

(5

Hed

s

Gordon, E. (1965). The Musical Aptitude Profile: a new and unique musical aptitude test battery. Bulletin of the Council for Research in Music Education, 6, 12-16.

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

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

Gordon, E. (1984). A longitudinal predictive validity study of the intermediate measures of music audiation. Bulletin of the Council for Research in Music Education, 78, 1-23.

Gordon, E. (1986). Final results of a two-year longitudinal predictive validity study of the Instrument Timbre Preference Test and the Musical Aptitude Profile. Bulletin of the Council for Research in Music Education, 89, 8-17.

Gregory, D. & Sims, W. L. (1987). Music preference analysis with computers. Journal of Music Therapy, 24 (4), 203-212.

Gronlund, N. E. (1981). Measurement and evaluation in teaching (4th ed.). New York: Macmillan.

Gronlund, N. E. (1985). Stating objectives for classroom instruction (3rd ed.). New York: Macmillan.

Harder, P. O. (1986). Basic materials in music theory (6th ed.). Boston : Allyn and Bacon.

Harder, P. O. (1985). Harmonic materials in tonal music : a programed course (5th ed.). Boston : Allyn and Bacon.

Heddon, S. K. (1982). Prediction of music achievement in the elementary school. Journal of Research in Music Education, 30, 61-68.

Helwig, C.

use of

Educat

Horst, P.

colleg

and T

Hufstade

throu

52-57

Humphr

audi

Edu

Johnsor

J. S.

Johnso

Bac

Karma

10,

Karma

Co

Krueg

te

30

Krue

s

Helwig, C. & Thomas, M. S. (1973). Predicting choral achievement through use of musicality and intelligence scores. Journal of Research in Music Education, 21, 276-280.

Horst, P. (1959). Relationship between preadmission variables and success in college. Seattle, WA: University of Washington Division of Counseling and Testing Services.

Hufstader, R. A. (1974). Predicting success in beginning instrumental music through use of selected tests. Journal of Research in Music Education, 22, 52-57.

Humphreys, J. T. (1986). Measurement, prediction, and training of harmonic audiation and performance skills. Journal of Research in Music Education, 34, 192-199.

Johnson, T. (1982). An analytical survey of the fifteen two-part inventions by J. S. Bach. Washington, D.C.: University Press of America.

Johnson, T. (1986). An analytical survey of the fifteen symphonias by J. S. Bach. Lanham, MD: University Press of America.

Karma, K. (1982). Validating tests of musical aptitude. Psychology of Music, 10, 33-35.

Karma, K. (1983). Selecting students to music instruction. Bulletin of the Council for Research in Music Education, 75, 23-32.

Krueger, R. J. (1972). A predictive investigation of personality and music teaching success. Bulletin of the Council for Research in Music Education, 30, 11-17.

Krueger, R. J. (1976). An investigation of personality and music teaching sucess. Bulletin of the Council for Research in Music Education, 47, 16-25.

Lee, R. E.
with co
in Mus
Lehman, I
NJ: P
Lehman,
Journ
Markle, S
& So
Marquis
Marshal
MA:
McArth
Jour
Meier,
and
Ev
Milste
co
M
Mour
b
C
C

Lee, R. E. (1967). An investigation of the use of the Musical Aptitude Profile with college and university freshman music students. Journal of Research in Music Education, 15, 278-288.

Lehman, P. R. (1968). Tests and measurements in music. Englewood Cliffs, NJ: Prentice-Hall.

Lehman, P. R. (1969). The predictive measurement of musical success. Journal of Research in Music Education, 17, 14-31.

Markle, S. M. (1969). Good frames and bad (2nd ed.). New York: John Wiley & Sons

Marquis' Who's Who in America. Chicago: Macmillan.

Marshall, J. C. & Hales, L. W. (1971). Classroom test construction. Reading, MA: Addison-Wesley.

McArthur, D. L. & Choppin, B. H. (1984). Computerized diagnostic testing. Journal of Educational Measurement, 21, 391-397.

Meier, S. T. & Geiger, S. M. (1986). Implications of computer-assisted testing and assessment for professional practice and training. Measurement and Evaluation in Counseling and Development, 19, 29-34.

Milstein, S. (1987). Acceptability and reliability of sensitive information collected via computer interview. Education and Psychological Measurement, 47, 523-533.

Mountford, R. D. (1982). Significant predictors of college band participation by college freshmen with high school band experience. Bulletin of the Council for Research in Music Education, 70, 87-89. (Reviewed by James E. Croft).

Neely, J.
aural
trainin
(Revi
Nunnally
McGr
Perry, W
profi
3996
Professi
Radocy
non
in M
Ruggie
Me
Ruggie
Pul
Ruggi
Ruggi
Mc
Samp
st
G
Schle
u
E

Neely, J. K. (1965). An evaluation of tests given to interrelate intelligence, aural acuity and musical achievement for purposes of prognosis in ear-training. Bulletin of the Council for Research in Music Education, 5, 51-54. (Reviewed by Paul Lehman).

Nunnally, J. (1961). Educational measurement and evaluation. New York: McGraw Hill.

Perry, W. W. (1965). A comparative study of selected tests for predicting proficiency in collegiate music theory. Dissertation Abstracts, 26, 3995-3996.

Professional Composer. (1988). Boston, MA: Mark of the Unicorn.

Radocy, R. E. (1972). Computerized criterion-referenced testing of certain nonperformance musical behaviors. Bulletin of the Council for Research in Music Education, 28, 1-6.

Ruggiero, C. H. (1989). Dances and other movements (musical score). Medfield, MA: Dorn Publications.

Ruggiero, C. H. (1989). Interplay (musical score). Medfield, MA: Dorn Publications.

Ruggiero, C. H. (1986). Set analysis programs. Okemos, MI: Author.

Ruggiero, C. H. (1982). Three blues for saxophone quartet (musical score). Medfield, MA: Dorn Publications.

Sampson, J. P. (1983). Computer-assisted testing and assessment: current status and implications for the future. Measurement and Evaluation in Guidance, 15, 293-299.

Schleuter, S. L. (1974). Use of standardized tests of musical aptitude with university freshmen music majors. Journal of Research in Music Education, 22, 258-269.

Schmitz, S. M. (1956). An investigation of the prognostic value of the revised Seashore tests and the Kwalwasser-Ruch Test of Musical Accomplishment for academic success in a music education program. Dissertation Abstracts, 17, 3042.

Swezey, R. W. (1981). Individual performance assessment: an approach to criterion-referenced test development. Reston, VA: Reston Publishing.

Taylor, E. M. (1941). A study in the prognosis of musical talent. Journal of Experimental Education, 10, 1-28.

Turner, G. A. (1987). Computers in testing and assessment: contemporary issues. International Journal of Instructional Media, 14, 187-197.

Turrentine, E. M. (1962). Predicting success in practice teaching in music. Bulletin of the Council for Research in Music Education, 4, 77-79. (Reviewed by G. H. Farwell).

Webber, G. H. (1976). The effectiveness of musical and nonmusical measures as predictors of success in beginning instrumental music classes. Bulletin of the Council for Research in Music Education, 49, 36-39. (Reviewed by Robert L. Cowden).

Wedman, J. F. & Stefanich, G. P. (1984). Guidelines for computer-based testing of student learning of concepts, principles, and procedures. Educational Technology, 24, 23-28.

Whellams, F. S. (1970). The relative efficiency of aural-musical and non-musical tests as predictors of achievement in instrumental music. Bulletin of the Council for Research in Music Education, 21, 15-21.

Whybrew, W. E. (1971). Measurement and evaluation in music. Dubuque, IA: Wm. C. Brown.

Whybrew, W. E. (1973). Research in evaluation in music education. Bulletin of the Council for Research in Music Education, 35, 9-17.

Willinghar

College

Young, W

musica

predic

Young, W

and u

Educ

Willingham, W. W. (1974). College placement and exemption. New York: College Entrance Examination Board.

Young, W. T. (1969). An investigation of the relative and combined power of musical aptitude, general intelligence and academic achievement tests to predict musical attainment. Dissertation Abstracts International, 30, 758A.

Young, W. T. (1972). Musical Aptitude Profile norms for use with college and university nonmusic majors. Journal of Research in Music Education, 20, 385-390.

MICHIGAN STATE UNIV. LIBRARIES
31293005728401