

*Project funded by the South Carolina State Department of Education
through contract with the Office of Program Evaluation, College of Education,
University of South Carolina*



**TECHNICAL DOCUMENTATION FOR THE
SOUTH CAROLINA ARTS ASSESSMENT PROGRAM (SCAAP):
ENTRY-LEVEL MUSIC & VISUAL ARTS ASSESSMENT 2009**

A report submitted to

R. Scot Hockman

*Education Associate for the Visual and Performing Arts
South Carolina State Department of Education*

Prepared by:

Robert Petrusis, Executive Director
Min Zhu, Research Assistant
Xiaofang Zhang, Research Assistant
Ashlee Lewis, Research Assistant

Office of Program Evaluation,
College of Education
University of South Carolina



TABLE OF CONTENTS

I.	INTRODUCTION	4
	SCAAP Purpose and Goals.....	4
	SCAAP History.....	4
	Dissemination & Research.....	5
	SCAAP Collaborators	6
II.	TEST FRAMEWORK & CONTENT	7
	SCAAP Format and Test Items.....	7
	Stimulus Material.....	7
	Item Review	7
	SCAAP Website.....	8
	Music Assessment Content & Test Specifications	8
	SC Music Content and Achievement Standards	8
	SCAAP Music Test Specifications	10
	Music Multiple-Choice Section: Format & Scoring.....	11
	Music Performance Task Section: Format & Scoring	12
	Visual Arts Content & Test Specifications	14
	SC Visual Arts Content and Achievement Standards.....	15
	SCAAP Visual Arts Test Specifications.....	16
	Visual Arts Multiple-Choice Section: Format & Scoring.....	16
	Visual Arts Performance Task Section: Format & Scoring.....	16
III.	TEST ADMINISTRATION	22
	Participants.....	22
	Training Test Administrators.....	22
	Administration Procedures.....	22
	Multiple-Choice Administration: Music & Visual Arts	23
	Performance Task Administration: Music	23
	Performance Task Administration: Visual Arts.....	23
IV.	SCORING	24
	Multiple-Choice Section	24
	Performance Task Section	24
	Benchmarking Music and Visual Arts Performance Tasks	24
	Web-based Rating System and Procedure	24
V.	ANALYSIS AND RESULTS.....	26
	Multiple-Choice Section	26
	Analysis.....	26
	Reliability.....	26
	Descriptive Statistics.....	28
	IRT Model Fit	29
	Equating Test Forms	29
	Technical Characteristics of Items.....	30
	Performance Task Section	35
	Inter-Rater Reliability of Performance Tasks	35
	Descriptive Statistics.....	37
	Task Difficulty Index and Discrimination Index	39
	Internal Validity	40

VI.	RESULTS BY SCHOOL.....	42
VII.	TEST ADMINISTRATOR FEEDBACK SURVEY RESULTS	47
	Training and Support	47
	Comments and Suggestions	47
	Recommendations and Clarifications	47
	Administering the SCAAP Assessments	48
	Multiple-Choice Assessment	48
	Comments and Suggestions	49
	Recommendations and Clarifications	49
	Music Performance Task Assessment	49
	Visual Arts Performance Task Assessment	50
	Using the SCAAP Assessment Results.....	50
	Additional Comments/Suggestions.....	51
VIII.	REFERENCES	53
IX.	APPENDICES	55
	APPENDIX A: 2009 SCAAP Item Analysis Results for Music Form 1	55
	APPENDIX B: 2009 SCAAP Item Analysis Results for Music Form 2.....	56
	APPENDIX C: 2009 SCAAP Item Analysis Results for Visual Arts Form 1	57
	APPENDIX D: 2009 SCAAP Item Analysis Results for Visual Arts Form 2	58
	APPENDIX E: 2009 SCAAP Item P-Values for Music Test Forms by Gender.....	59
	APPENDIX F: 2009 SCAAP Item P-Values for Music Test Forms by Ethnicity	60
	APPENDIX G: 2009 SCAAP Item P-Values for Visual Arts Test Forms by Gender	61
	APPENDIX H: 2009 SCAAP Item P-Values for Visual Arts Forms by Ethnicity	62

LIST OF FIGURES AND TABLES

<i>Figure 5.1. P-value Histograms for Music and Visual Arts Test Forms</i>	30
Table 1.1 <i>List of SCAAP Advisory Committee Members Music and Visual Arts</i>	6
Table 2.1 <i>Table of Specifications for SCAAP Music Assessment</i>	11
Table 2.2 <i>Number of Items in Each Part of the Music Multiple-Choice Test Forms</i>	11
Table 2.3 <i>SCAAP Music Task 1 Rubric—Tonal Criteria</i>	12
Table 2.4 <i>SCAAP Music Task 1 Rubric—Rhythm Criteria</i>	13
Table 2.5 <i>SCAAP Music Task 1 Rubric—Vocal Quality Criteria</i>	13
Table 2.6 <i>SCAAP Music Task 2 Rubric—Rhythm Criteria</i>	13
Table 2.7 <i>SCAAP Music Task 2 Rubric—Improvisation Criteria</i>	14
Table 2.8 <i>Maximum Points for Music Assessment Performance Tasks</i>	14
Table 2.9 <i>Table of Specifications for SCAAP Visual Arts Assessment</i>	16
Table 2.10 <i>SCAAP Visual Arts Task 1 Rubric—Compare/Contrast</i>	18
Table 2.11 <i>SCAAP Visual Arts Task 2a Rubric—Drawing</i>	19
Table 2.12 <i>SCAAP Visual Arts Task 2b Rubric—Self-Critique</i>	20
Table 2.13 <i>Maximum Points for Visual Arts Performance Tasks</i>	21
Table 5.1 <i>Reliability Indices for SCAAP 2008 Music and Visual Arts Multiple-Choice Items</i> .	26
Table 5.2 <i>Reliability of “Understanding and Listening” for Music Forms 1 and 2</i>	27
Table 5.3 <i>Reliability Indices for Music Multiple-Choice Assessment by SC Standard</i>	27
Table 5.4 <i>Reliability Indices for Visual Arts Multiple-Choice Assessment by SC Standard</i>	28
Table 5.5 <i>Descriptive Statistics for Music and Visual Arts Test Forms</i>	28
Table 5.6 <i>Ranges of P-values and Discrimination Indices of Test Forms</i>	31
Table 5.7 <i>Number of Items with Low and Negative Discrimination Indices</i>	32
Table 5.8 <i>Descriptive Statistics of Test Scores by Gender</i>	32
Table 5.9 <i>Descriptive Statistics of Test Scores by Ethnicity</i>	33
Table 5.10 <i>Ranges of P-values of Test Forms by Gender and Ethnic Groups</i>	33
Table 5.11 <i>Summary of DIF Classification for Multiple-Choice Test Forms</i>	35
Table 5.12 <i>Inter-Rater Reliability of Music Performance Tasks Using Analytic Rubrics</i>	36
Table 5.13 <i>Inter-Rater Reliability of Visual Arts Performance Tasks Using Holistic Rubrics</i> ...	36
Table 5.14 <i>Descriptive Statistics for Music Performance Tasks Using Analytic Rubrics</i>	38
Table 5.15 <i>Descriptive Statistics for Visual Arts Performance Tasks Using Holistic Rubrics</i> ...	39
Table 5.16 <i>Task Difficulty Indices for Music and Visual Arts Performance Tasks</i>	39
Table 5.17 <i>Discrimination Index for Music and Visual Arts Performance Tasks</i>	40
Table 5.18 <i>Correlations for Music Multiple-Choice Test Forms and Performance Tasks</i>	40
Table 5.19 <i>Correlations for Visual Arts Multiple-Choice Test Forms and Performance Tasks</i> .	41
Table 6.1 <i>Music Assessment Results for 2008 and 2009 by School</i>	43
Table 6.2 <i>Visual Arts Assessment Results for 2008 and 2009 by School</i>	45

I. INTRODUCTION

SCAAP Purpose and Goals

Established in 2000, the South Carolina Arts Assessment Program (SCAAP) is a collaborative effort among the South Carolina Department of Education (SCDE), the Office of Program Evaluation (OPE) at the University of South Carolina (USC), and South Carolina arts educators. The purpose of SCAAP is to provide arts educators and school administrators with a tool to authentically measure their students' arts achievement and to objectively evaluate their schools' arts programs based on the South Carolina Visual and Performing Arts Curriculum Standards from 2003 (SCDE, 2003). The goal of SCAAP is to develop separate standards-based arts assessments in dance, music, theatre, and visual arts that are available for use at several K-12 grade levels. Currently, SCAAP has four entry-level and two middle-level arts assessments, each with an online multiple-choice section and a performance tasks section. All SCAAP assessment items are developed by South Carolina arts educators and evaluated by measurement specialists at OPE.

SCAAP History

In 2000 (Year 1), the project began with a focus on developing and field testing large-scale assessments for elementary school music and visual arts. Advisory Committees, comprised of statewide leaders in music and visual arts education, were formed to determine the test content and format. Advisory Committee members met with the SCAAP personnel to determine test specifications by selecting the content and achievement standards most appropriate for large-scale assessment. The test specifications provided the framework from which the committee members selected the most appropriate arts test population and format for the assessments.

Committee members chose elementary students as the initial test population for the music and visual arts assessments because a majority of South Carolina elementary schools offered music and visual arts programs. The committee members chose 4th grade as the elementary grade level to be assessed because that grade level is included South Carolina music and visual arts curriculum for grades three through five. Additionally, the committee members felt that teachers would not have had enough instructional time to cover those standards for students to be tested in the 3rd grade and would not have enough time to use the assessment results to modify instruction if necessary in the 5th grade.

The item formats chosen were multiple-choice and on-demand performance tasks. Committee members determined that those two formats were most suited to assess students' music and visual arts achievement in relation to the 2003 *SC Visual and Performing Arts Curriculum Standards*. The Advisory Committee members then created test items and tasks to match the test specifications. The SCAAP personnel used those items and tasks to assemble and field-test three 40-item multiple-choice test forms and two performance tasks each for music and for visual arts. Statistical analyses were conducted to evaluate the quality of the assessments and participating teachers were surveyed to gather feedback regarding test administration. A detailed description of the Year 1 results and activities can be found in the report *Technical Documentation for the South Carolina Arts Assessment Project (SCAAP) Year 1: Fourth Grade Music and Visual Arts Assessments* submitted to the SCDE.

In Year 2, the SCAAP 4th grade music and visual arts assessments were refined and pilot-tested. Based on statistical analyses from Year 1, the SCAAP personnel assembled two, rather than three, multiple-choice test forms and increased the number of items on each test to 45 to obtain adequate reliability. In Year 2, the SCAAP personnel also pilot-tested the web-based music and visual arts assessment prototype. A detailed description of SCAAP activities conducted during Year 2 can be found in the report *Technical Documentation for the South Carolina Arts Assessment Project (SCAAP) Year 2: Fourth Grade Music and Visual Arts Assessments* submitted to the SCDE. In Year 3, the SCAAP personnel began implementing the web-based SCAAP 4th grade music and visual arts assessments at several schools across the state in conjunction with the Distinguished Arts Program (DAP) grants awarded by the SCDE.

Advisory Committees were formed in Year 3 to begin developing the entry-level SCAAP dance and theatre assessments. In Year 4, the web-based multiple-choice sections of the entry-level dance and theatre assessments were field-tested at several middle and high schools across South Carolina. In Year 5, the SCAAP personnel worked with the Advisory Committee members to develop performance tasks for dance and theatre; both sections of the entry-level SCAAP dance and theatre assessments (i.e., multiple-choice and performance tasks) were field-tested in Year 6 and implemented at several schools across the state in Year 7. Also in Year 6, Advisory Committees were established to begin development of the SCAAP music and visual arts assessments for middle school students. The SCAAP middle school music and visual arts assessments were field-tested in Year 7. In Year 7, SCAAP comprised six different assessments—four entry-level and two intermediate-level assessments. All SCAAP assessments include a web-based multiple-choice section and a performance tasks section. Of the six SCAAP assessments, five include two performance tasks and one assessment includes three performance tasks. In the current year, Year 8, only the entry-level music and visual arts assessments were administered. Both assessments were administered during the spring semester, and school-level results were reported to teachers and principals the following fall semester.

Dissemination & Research

As the only web-based and fully-implemented arts assessment in the country, the SCAAP has been showcased at state and national conferences, such as the South Carolina Alliance for Arts Education (SCAAE), South Carolina Educators for the Practical Use of Research (SCEPUR), Music Educators National Conference (MENC), American Evaluation Association (AEA), and American Educational Research Association (AERA). The SCAAP personnel have presented on topics ranging from development and implementation of the assessments to validation of remote rating procedures. In addition, the SCAAP has been highlighted in several publications, including an assessment textbook, *Assessing Performance: Designing, Scoring, and Validating Performance Tasks* (Johnson, Penny, & Gordon, 2008), and *Assessment in Music Education: Integrating Curriculum, Theory, and Practice* from the Proceedings of the 2007 Symposium on Assessment in Music Education (Yap & Pearsall, 2007). Contact the SCAAP personnel at scaap@mailbox.sc.edu for a complete list of presentations, publications, and research utilizing the SCAAP.

SCAAP Collaborators

Because of the collaborative nature of SCAAP, several organizations and individuals are continuously involved in the development, implementation, and maintenance of the program.

The SCAAP personnel, consisting of a team of graduate students in the OPE led by Dr. Robert Petrulis, perform tasks related to the educational measurement aspects of the program as well as the administrative aspects of the program. Previously, Dr. Ching Ching Yap led the program. Currently, Dr. Petrulis and the SCAAP personnel advise and support members of all SCAAP Advisory Committees and also work with staff members from Enterprise Applications, formerly eBusiness Solutions, at USC to maintain and continually update the SCAAP web-based assessments.

The South Carolina Department of Education (SCDE) serves as the funding agency for SCAAP, and R. Scot Hockman, Education Associate for the Visual and Performing Arts at the SCDE, provides guidance for the future of the program as the SCDE representative. Mr. Hockman actively participates in many of the sessions involving the SCAAP Advisory Committees.

The SCAAP Advisory Committees are comprised of educators acknowledged statewide as leaders in their respective arts areas and who have completed the Curriculum Leadership Institute of the Arts (CLIA) and Arts Assessment Institute (AAI). Committee members guide the assessments by providing content-area expertise at several points throughout the year. During Item Review sessions, committee members create new multiple-choice test items, edit and refine existing test items, and review and refine performance task documents. During Validation Sessions, the Advisory Committee members review and refine performance task documents and benchmark student performance tasks for subsequent rating. Table 1.1 is a list of the core members of the SCAAP Advisory Committee for the 4th grade music and visual arts assessments.

Table 1.1

List of SCAAP Advisory Committee Members Music and Visual Arts

SCAAP Advisory Committee Members	
Music	Visual Arts
Kathy Clark	Connie Boleman
Pam Gowan	Stan Dubose
Mark Hodges	Lillie Dunning
Heather Turner	Laura McFadden
Tara Pearsall	

II. TEST FRAMEWORK & CONTENT

SCAAP Format and Test Items

The SCAAP test format, developed in Year 1 and modified slightly in Year 2, includes two 45-item parallel multiple-choice test forms and two performance tasks. The format is the same for both the music and the visual arts assessment. All test items were developed by members of the music and visual arts Advisory Committees along with arts faculty members from universities and colleges across South Carolina. Each multiple-choice test item used in the SCAAP assessment targets a South Carolina content and achievement standard and has a designated Bloom's taxonomy level. All multiple-choice items have four options, and many include multimedia interpretive material. Each performance task targets a specific South Carolina content standard and has a standardized administration procedure.

Stimulus Material

Stimulus materials used in SCAAP multiple-choice test items include multimedia interpretive materials such as artwork reprints and music notation images. The file formats include mp3 and wav (audio files), jpeg and gif (image files), and flv and swf (video files). Often, the music examples and images used in the SCAAP test items were created by local South Carolina artists and composers.

Item Review

Many of the multiple-choice items used in the SCAAP assessments from 2004 to present were created during the initial development phase of the assessment; however, each year new items are created to augment the item banks and to replace items terminated due to poor item performance. All items are reviewed and edited yearly by Advisory Committee members and the SCAAP personnel based on item analysis from the previous year's assessment results. The committee members and the SCAAP personnel also review test items for (a) age-appropriateness and readability, (b) alignment to state curriculum standards, (c) gender and ethnicity bias using Differential Item Functioning (DIF) analysis.

Any changes to existing test items are determined by the committee members during the item review sessions and submitted to the SCAAP website following the sessions. Any changes not approved during the session are sent to committee members for review before test administration. The SCAAP personnel are responsible for the final appearance and/or sound of all test items and stimulus material. In addition, the SCAAP personnel make sure that all test items adhere to the following item writing guidelines.

Item-Writing Guidelines

1. The item relates directly to a specific standard.
2. The item requires students to use higher-order thinking skills.
3. The stem is a complete question or an incomplete statement; wording is simple and clear.
4. Information in the stem does not cue the answer.
5. Negative stems are avoided.
6. There is only one clear correct answer.
7. The correct answer is varied and options are arranged in a logical order (i.e., "abc").
8. Each alternative is plausible to a student who lacks the targeted knowledge.
9. Overlapping alternatives are avoided; each option is independent and mutually exclusive.

10. Alternatives are parallel in concept, language structure, grammar, and appearance.
11. Item options are equal or nearly equal in length.
12. Options avoid repeated words that are better suited in the stem.
13. Language usage and grammar in the stem and options are correct.
14. Wording in the stem and options are simple and clear.
15. The use of “all of the above” and “none of the above” as options is avoided.

SCAAP Website

Beginning in Year 3, all SCAAP multiple-choice assessments were administered online via the SCAAP website (<https://scaap.ed.sc.edu>). In addition, other aspects of the assessment became web-based such as performance task rating and monitoring. The SCAAP website is secured by usernames and passwords assigned by the SCAAP personnel who serve as the website administrators. Some of the website administrative responsibilities include (a) uploading audio, image, and video stimulus material, (b) creating and revising multiple-choice test items, (c) assembling multiple-choice test forms from the item database, (d) monitoring the online remote rating system. Depending on their access level, website users can perform such activities as (a) registering students to take the assessments, (b) uploading performance tasks files, (c) benchmarking student performance tasks for rater training purposes, (d) completing rater training and rating live student performance tasks, and (e) viewing school arts assessment results.

Due to the complexity of the website, the SCAAP personnel have documented the various features and functions in a website manual. The SCAAP personnel have also created separate manuals for each user access level to help facilitate website usage. The SCAAP website uses a SQL-DB server to store the large amount of data needed to conduct the SCAAP assessment and to accommodate the large volume of concurrent users that occurs during SCAAP test administration and performance task rating.

In Year 8, a blog (<http://www.scaapblog.blogspot.com>) for test administrators was set up in addition to the SCAAP website. The blog provides supportive information to test administrators, such as videos demonstrating how to record sound files and how to use features of the SCAAP website, as well as information regarding when and how to access SCAAP test results. The blog was maintained and updated by SCAAP personnel on a regular basis.

Music Assessment Content & Test Specifications

The 4th grade music assessment was developed to assess students’ music achievement based on the *South Carolina Visual and Performing Arts Curriculum Standards* (2003) for grades three through five. With the SCAAP personnel, the SCAAP Advisory Committee members selected music content and achievement standards appropriate for large-scale assessment and constructed a table of test specifications. The SCAAP Music Assessment includes two sections: multiple-choice items and performance tasks. Multiple-choice items require students to demonstrate their knowledge of and skills in (a) music vocabulary, (b) notation, (c) listening music, (d) evaluation of performance problems, and (e) performance of music skills. Performance tasks require students to demonstrate their singing and improvisation skills on demand.

SC Music Content and Achievement Standards

Following are the music content and achievement standards selected by the Music Advisory Committee for inclusion in the SCAAP 4th grade music assessment.

Content Standard 1: Singing, alone and with others, a varied repertoire of music

Achievement Standards: Students will:

- a. Sing independently, on pitch, and in rhythm, using appropriate timbre, diction, and posture while maintaining a steady tempo
- b. Sing expressively, alone or in groups, blending vocal timbres, matching dynamic levels, and responding to the cues of a conductor
- c. Sing, alone and with others, a varied repertoire of music including partner songs, descants, ostinati, and rounds

Content Standard 3: Improvising melodies, variations, and accompaniments

Achievement Standards: Students will:

- a. Improvise, in the same style, responses to given rhythmic and melodic patterns
- b. Improvise simple rhythmic and melodic ostinato patterns and accompaniments
- c. Improvise simple rhythmic variations and melodic embellishments
- d. Improvise short songs and instrumental pieces using traditional and nontraditional sound sources

Content Standard 4: Composing and arranging music within specified guidelines

Achievement Standards: Students will:

- a. Compose and arrange music using standard and nonstandard notation.
- b. Compose and arrange music to accompany readings and dramatizations.
- c. Compose and arrange short songs and instrumental pieces within specific guidelines, using basic music elements.
- d. Compose and arrange using a variety of sound sources

Content Standard 5: Reading and notating music

Achievement Standards: Students will:

- a. Read and write rhythmic notation incorporating syncopation as well as whole, half, quarter, eighth, and sixteenth notes and corresponding rests.
- b. Read and write short melodic notation in pentatonic, major, and minor tonalities.
- c. Identify symbols and terminology for dynamics, tempo, and articulation and interpret them correctly when performing.
- d. Write notation using standard symbols for meter, rhythm, pitch, and dynamics.

Content Standard 6: Listening to, analyzing, and describing music

Achievement Standards: Students will:

- a. Identify examples of music forms including motive to phrase, 4-bar phrase, canon, rondo, AABA, 12-bar blues, and theme and variation.
- b. Demonstrate perceptual skills by moving, answering questions, and describing selections representing diverse musical styles.
- c. Use appropriate terminology to explain pitch, notation, meter, chords, voices, instruments, and performances.
- d. Explain music using the appropriate terminology for pitch, notation, meter, chords, voices, instruments, and performances.
- e. Identify by sight and sound a variety of instruments including orchestral, band, multicultural, and digital.
- f. Demonstrate movement and emotional response to prominent music characteristics while listening.

- g. Identify music in pentatonic, major, and minor tonalities.

Content Standard 7: Evaluating music and music performances

Achievement Standards: Students will:

- a. Devise criteria for evaluating performances and compositions based upon musical concepts, ideas, and values.
- b. Use appropriate music terminology to explain their personal preferences for specific musical works and styles.
- c. Apply music concepts when judging the quality of their own performances and those of others and when offering constructive suggestions for improvement.

Content Standard 8: Understanding relationships between music, the other arts, and disciplines outside the arts

Achievement Standards: Students will:

- a. Explain the role of music in life experiences, celebrations, community functions, and special events.
- b. Identify similarities and differences in the meanings of common terms used in the various arts disciplines (e.g., “texture,” “color,” “form”).
- c. Explain how the principles and subject matter of disciplines outside the arts interrelate with those of music.

Content Standard 9: Understanding music in relation to history and culture

Achievement Standards: Students will:

- a. Listen to examples of music from various historical periods and world cultures and identify the pieces by genre or style.
- b. Describe how elements of music are used in music examples from various cultures of the world.
- c. Identify various uses of music in daily experiences and describe the characteristics that make a particular type of music suitable for each use.
- d. Identify and describe the roles of musicians in various settings and world cultures.
- e. Demonstrate audience behavior appropriate for the context and style of music being performed.

SCAAP Music Test Specifications

Table 2.1 presents the test specifications for the music assessments. The table presents the percentages of items assessing each content standard for the multiple-choice section of the assessment and the content standard addressed by each performance task. The committee members decided to exclude *SC Content Standard 2: Playing instruments alone and with others* from the SCAAP 4th grade assessment due to concerns of equitable access to instruments across the state.

Table 2.1
Table of Specifications for SCAAP Music Assessment

Content Standard	Percentages
1. Singing	Performance Task 1
3. Improvisation	Performance Task 2
4. Composition	10%
5. Reading and Notating	30%
6. Analysis	25%
7. Evaluation	15%
8. Connections	5%
9. History and Culture	15%

Music Multiple-Choice Section: Format & Scoring

In Year 1, the SCAAP multiple-choice section included 40 multiple-choice items and two performance tasks for assessing fourth-grade students’ music achievement. Based on analysis of the Year 1 results, the SCAAP personnel recommended increasing the number of test items in the multiple-choice section from 40 to 45 items to achieve satisfactory test reliability. Beginning in Year 2, the multiple-choice section of the music assessment consisted of two parallel test forms, each with 45 multiple-choice items, 25 of which were repeated in both test forms. Each test form was divided into two parts: “Understanding Music” (approximately 30 items) and “Listening to Music” (approximately 15 items).

Student responses to the 45-item, web-based multiple-choice test forms are stored on the SCCAP website and scored automatically. Each correct answer is scored as 1 and each incorrect answer is scored as 0. The maximum score for the SCAAP multiple-choice section is 45 points. Table 2.2 shows the total points possible for the music multiple-choice section as well as the number of items included in each part of each multiple-choice test form for Year 8.

Table 2.2
Number of Items in Each Part of the Music Multiple-Choice Test Forms

Parts	Music Form 1	Music Form 2
Understanding Music	29	29
Listening to Music	16	16
Total	45	45

Music Performance Task Section: Format & Scoring

In Year 1, the Music Advisory Committee and SCAAP personnel developed two music performance tasks, which were the same tasks used in Year 8. SCAAP Music Performance Task 1 requires students to individually perform a familiar song on a neutral syllable (“du”). Music Performance Task 2 requires students to individually perform an 8-beat rhythm improvisation using rhythm syllables. The student directions for both tasks are recorded on a compact disc (CD) to standardize test administration. Each test administrator is required to play the CD directions for each student and then digitally record that student performing the task when prompted.

Beginning in Year 3, the SCAAP music performance tasks are scored by trained raters using hierarchical analytic rubrics¹. After comparing the inter-rater reliability obtained using holistic and analytic rubrics in Year 2, Music Advisory Committee members decided to use hierarchical analytics rubrics because they would (a) allow raters to identify and evaluate the different components of the students’ performances separately, and (b) provide participating music teachers with detailed feedback regarding students’ strengths and weaknesses. The rubric for Music Performance Task 1 includes three criteria (Tonal, Rhythm, and Vocal Quality) and the rubric for Music Performance Task 2 includes two criteria (Rhythm and Improvisation). Each criterion has 5 levels, ranging from 0 to 4, and each level represents a skill to be accomplished. The order of the levels is based on the hierarchy of skill development. The rubrics used to score the Year 8 music performance tasks are presented in Tables 2.3 to 2.7. Since Year 7, the music performance tasks rubrics have been provided to test administrators prior to testing.

Table 2.3
SCAAP Music Task 1 Rubric—Tonal Criteria

Rating	Tonal Criteria
4	Intonation and pitches are accurate
3	Tonal center is established and maintained
2	Major Tonality is established
1	Melodic contour is accurate
0	Incorrect melodic contour; incomplete performance or performance with pause(s) or stops

¹ The analytic rubrics used to score the SCAAP performance tasks were developed based on Gordon’s “Rating Scales and Their Uses for Measuring and Evaluating Achievement in Music Performance” (2002). According to Gordon, analytic rubrics, or continuous rating scales as he refers to them, are used to measure each dimension of a performance (e.g., rhythm component or tonal component).

Table 2.4
SCAAP Music Task 1 Rubric—Rhythm Criteria

Rating	Rhythm Criteria
4	Tempo is consistent, and rhythm patterns are accurate
3	Tempo and meter are maintained nearly all the time; Rhythm patterns are mostly correct
2	Duple meter is established
1	A tempo is established
0	Incomplete performance or tempo is not established

Table 2.5
SCAAP Music Task 1 Rubric—Vocal Quality Criteria

Rating	Vocal Quality Criteria
4	Consistent use of head voice/singing voice and consistent breath support
3	Consistent use of head voice/singing voice BUT minimal breath support
2	Minimal use of head voice/singing voice
1	No use of head voice/singing voice
0	Incomplete performance

Table 2.6
SCAAP Music Task 2 Rubric—Rhythm Criteria

Rating	Rhythm Criteria
4	Macrobeats and microbeats, divisions, or elongations are accurately represented
3	A tempo is maintained most of the time, and meter is clearly defined
2	Meter is established
1	A tempo is established
0	Incomplete performance or no tempo established

Table 2.7
SCAAP Music Task 2 Rubric—Improvisation Criteria

Rating	Improvisation Criteria
4	Improvised a rhythm pattern using complex rhythm patterns such as divisions or elongations
3	Tempo and meter are maintained nearly all the time; Rhythm patterns are mostly correct
2	Duple meter is established
1	A tempo is established
0	Incomplete performance, no improvisation, or no recognizable improvisational intent

From Year 7, raters had the option of choosing augmentation scores (+ and -) for each student performance. With augmentation scores, the maximum score for each performance criteria for the SCAAP Music Performance Tasks is 4.33. For Music Task 1, the maximum score is 12.99 and for Music Task 2, the maximum score is 8.66. Table 2.8 presents a summary of the maximum points for each criterion and for each task for the SCAAP Music Performance Tasks.

Table 2.8
Maximum Points for Music Assessment Performance Tasks

Performance Task	Criteria	Points per Criteria	Points per Task
Task 1: Singing	Tonal	4.33	12.99
	Rhythm	4.33	
	Vocal Quality	4.33	
Task 2: Rhythm Improvisation	Rhythm	4.33	8.66
	Improvisation	4.33	

Visual Arts Content & Test Specifications

The 4th grade visual arts assessment was developed to assess students' visual arts achievement based on the *South Carolina Visual and Performing Arts Curriculum Standards* (2003) for grades three through five. Along with the SCAAP personnel, the SCAAP Advisory Committee members selected visual arts content and achievement standards appropriate for large-scale assessment and constructed a table of test specifications. The SCAAP Visual Arts Assessment includes two sections: multiple-choice items and performance tasks. Multiple-choice items

require students to demonstrate their knowledge of and skills related to (a) visual arts vocabulary, (b) the creation of artworks, (c) the evaluation of artistic choices, and (d) the understanding of visual arts in cultural and historical contexts. Performance tasks required students to demonstrate on demand their ability to compare and contrast artworks and to complete and critique their own artwork.

SC Visual Arts Content and Achievement Standards

Following are the visual arts content and achievement standards selected by the Visual Arts Advisory Committee for inclusion in the SCAAP 4th grade visual arts assessment. The visual arts content and achievement standards were selected from *South Carolina Visual and Performing Arts Curriculum Standards* from 2003 for grades three through five.

Content Standard 1: Understanding and applying media, techniques, and processes

Achievement Standards: Students will

- a. Describe how different media, techniques, and processes evoke different responses in the viewer of an artwork.
- b. Use a variety of media, techniques, and processes to communicate ideas, experiences, and stories through their artworks.
- c. Use art materials and tools in a safe and responsible manner.

Content Standard 2: Using knowledge of structures and functions

Achievement Standards: Students will

- a. Describe, both orally and in writing, how the various elements and principles of design function to evoke different responses in the viewer of an artwork.
- b. Select and use various elements and principles of design to communicate personal ideas in their artworks.

Content Standard 3: Choosing and evaluating a range of subject matter, symbols, and ideas

Achievement Standards: Students will

- a. Select and use subject matter, symbols, and ideas to communicate meaning through their artworks.
- b. Evaluate how particular choices of subject matter, symbols, and ideas function to communicate meaning in their own artworks and those of others.

Content Standard 4: Understanding the visual arts in relation to history and cultures

Achievement Standards: Students will

- a. Identify specific artworks and styles as belonging to particular artists, cultures, periods, and places.
- b. Identify a variety of artworks, artists, and visual arts materials that exist in South Carolina.
- c. Describe how history, culture, and the visual arts can influence one another.

Content Standard 5: Reflecting upon and assessing the merits of their work and the work of others

Achievement Standards: Students will

- a. Describe how an artist's experiences can influence the development of his/her artworks.
- b. Analyze their own artworks and those of others and describe improvements that could be made.
- c. Distinguish between personal preference and the objective analysis of artworks.

Content Standard 6: Making connections between visual arts and other disciplines

Achievement Standards: Students will

- a. Compare and contrast characteristics of the visual arts and other arts disciplines.
- b. Identify connections among the visual arts, other arts disciplines, and content areas across the curriculum.
- c. Recognize career opportunities in the visual arts.

SCAAP Visual Arts Test Specifications

Table 2.9 presents the test specifications for the SCAAP visual arts assessments. The table presents the emphasis placed on each content standard for both the multiple-choice and performance task sections of the SCAAP Visual Arts assessment combined.

Table 2.9

Table of Specifications for SCAAP Visual Arts Assessment

Content Standard	Emphasis
1. Applying Media	15%
2. Knowledge of Structures	25%
3. Evaluating a Range	15%
4. Understanding Visual Arts	15%
5. Assessing the Merits	20%
6. Making Connections	10%

Visual Arts Multiple-Choice Section: Format & Scoring

In Year 1, the SCAAP multiple-choice section included 40 multiple-choice items and two performance tasks for assessing fourth-grade students' visual arts achievement. Based on analysis of the Year 1 results, the SCAAP personnel recommended increasing the number of test items in the multiple-choice section from 40 to 45 items to achieve satisfactory test reliability. Beginning in Year 2, the multiple-choice section of the music assessment consisted of two parallel test forms, each with 45 multiple-choice items; there were 25 linking items between the two test forms. Student responses to the 45-item, web-based multiple-choice test forms are stored on the SCAAP website and scored automatically. Each correct answer is scored as 1 and each incorrect answer item is scored as 0. The maximum score for the SCAAP multiple-choice section is 45 points.

Visual Arts Performance Task Section: Format & Scoring

In Year 1, the Visual Arts Advisory Committee and SCAAP personnel developed two visual arts performance tasks, which were the similar tasks used in Year 7. SCAAP Visual Arts Performance Task 1 requires students to compare and contrast, in writing, two artworks using a word bank of visual arts terms. Visual Arts Performance Task 2 is a two-part task. Part one, Task 2a, requires students to complete a drawing based on a given prompt and part two (Task 2b) requires students to write a critique of their own drawing using a word bank of visual arts terms.

To standardize test administration, each student receives the same performance task booklet, which includes written directions and space to complete the tasks. Test administrators are responsible for making sure that students use # 2 pencils and work independently. Beginning this year (Year 8), each Visual Arts test administrator was provided with a Test Administrator Script to ensure that the delivery of directions was standardized for each student.

Since Year 1, the SCAAP visual arts performance tasks have been scored by trained raters using holistic rubrics. The SCAAP Visual Arts Validation Committee members decided to use holistic rubrics to describe students' performance levels because the criteria used to evaluate students' performance tasks are dependent on one another and difficult to separate when scoring. Each rubric has 5 levels, ranging from 0 to 4, and each level describes a student's achieved level of proficiency with regard to the specific performance task. The rubrics used to score this year's (Year 8) visual arts performance tasks are presented in Tables 2.10 through 2.12. Beginning this year, the rubrics were made available to test administrators prior to administration in both hard copy and on the SCAAP website.

Table 2.10
SCAAP Visual Arts Task 1 Rubric—Compare/Contrast

Rating	The writing should be characterized by most of the following:
4	<p>Most of the student’s writing demonstrates a clear understanding of the similarities and differences between the two artworks. Student uses at least 4 terms correctly when referring to the artworks, and clearly shows in the context of the writing that he/she understands the art terms used.</p> <p>Three of the four terms used are specific references. At least two of the four terms used are explanations. Only one additional term may be used incorrectly.</p>
3	<p>Most of the student’s writing demonstrates a clear understanding of the similarities and differences between the two artworks. Student uses at least 4 terms correctly when referring to the artworks, and clearly shows in the context of the writing that he/she understands the art terms used.</p> <p>At least two of the four terms used are specific references. At least one of the four of the terms used is an explanation.</p>
2	<p>Some of the student’s writing demonstrates a clear understanding of the similarities and/or differences between the two artworks. Student uses at least 3 terms correctly when referring to the artworks, and clearly shows in the context of the writing that he/she understands the art terms used.</p> <p>All three terms used are general or specific references. The terms may or may not include explanations.</p>
1	<p>Student’s writing demonstrates limited understanding of the similarities and/or differences between the two artworks. Student uses at least 2 terms correctly when referring to the artworks, and clearly shows in the context of the writing that he/she understands the art terms used.</p> <p>The two terms used may not include any specific references or explanations. The general references must demonstrate that the student understands the art term.</p>
0	<p>Student uses at most 1 term correctly and clearly shows in the context of the writing that he/she understands the art term used.</p> <p>Most of the writing is off topic or there is an insufficient amount of original writing to evaluate student’s visual arts knowledge.</p>

Table 2.11
SCAAP Visual Arts Task 2a Rubric—Drawing

Rating The drawing should be characterized by most of the following:

- 4 Drawing includes the assigned topic and fills the space in a **unified composition**. The drawing includes an environment with a clear distinction between the background and foreground creating a sense of depth in the whole composition. Drawing also includes **many** details. Texture and patterns are used to enhance the picture.
- 3 Drawing includes the assigned topic and fills the space but may or may not be unified. The drawing includes an environment with a clear distinction between the background and foreground that may or may not create a sense of depth in the whole composition. The drawing includes **some** details, texture, or patterns.
- 2 Drawing includes the assigned topic and fills most of the space. An attempt was made to include an environment and create depth, but there is not a clear distinction between background and foreground. Drawing includes **few** details, texture, or patterns.
- 1 Drawing includes the assigned topic and an attempt was made to fill the space. No attempt was made to include a background and foreground and to create a sense of depth in the composition. Drawing includes **very limited** details, texture, or pattern.
- 0 Drawing does not address assigned topic, or drawing does not attempt to fill the space, or drawings are unrecognizable.
-

Table 2.12
SCAAP Visual Arts Task 2b Rubric—Self-Critique

Rating	The writing should be characterized by most of the following:
4	<p>Most of the student’s writing demonstrates a clear understanding of the strengths and weaknesses of the drawing.</p> <p>Student uses at least 4 terms correctly when referring to his/her artwork, and clearly shows in the context of the writing that he/she understands the art terms used. All 4 terms are specific references. At least 2 of the 4 terms are explanations. Only one additional term may be used incorrectly.</p>
3	<p>Most of the student’s writing demonstrates a clear understanding of the strengths and weaknesses of the drawing.</p> <p>Student uses at least 4 terms correctly when referring to his/her artwork, and clearly shows in the context of the writing that he/she understands the art terms used. All 4 terms are specific references. At least 1 of the 4 terms is an explanation.</p>
2	<p>Some of the student’s writing demonstrates a clear understanding of the strengths and weaknesses of the drawing.</p> <p>Student uses at least 3 terms correctly when referring to his/her artwork, and clearly shows in the context of the writing that he/she understands the art terms used. All 3 terms are specific references. The terms used may or may not include explanations.</p>
1	<p>Student’s writing demonstrates limited understanding of the strengths and weaknesses of the drawing.</p> <p>Student uses at least 2 terms correctly when referring to his/her artwork, and clearly shows in the context of the writing that he/she understands the art terms used. All 2 terms are specific references. The terms used may or may not include explanations.</p>
0	<p>Student uses at most 1 term correctly, and clearly shows in the context of the writing that he/she understands the art term used.</p> <p>Most of the writing is off topic, or insufficient amount of original writing to evaluate the student’s visual arts knowledge.</p>

Beginning in Year 7, raters had the option of using augmentation scores (+ and -) when scoring students’ performances. Therefore, the maximum score for Visual Arts Task 1 is 4.33 and the maximum for Visual Arts Task 2 is 8.66 (4.33 for Task 2a and 4.33 for Task 2b). Table 2.13 presents a summary of the maximum points for each task of the SCAAP Visual Arts Performance Tasks.

Table 2.13
Maximum Points for Visual Arts Performance Tasks

Performance Task	Description	Points per Criteria	Points per Task
Task 1	Compare/Contrast	4.33	4.33
Task 2	Drawing	4.33	8.66
	Self-Critique	4.33	

III. TEST ADMINISTRATION

Participants

In Year 8, 24 schools and 8 school districts received Distinguished Arts Program (DAP) grants from the South Carolina State Department of Education (SCDE). All 24 schools that received school-level DAP grants were required to participate in the SCAAP 4th grade music and visual arts assessments. Each of the 8 school districts that received a district-level DAP grant registered three elementary schools to participate in the SCAAP 4th grade assessments; districts with three or fewer elementary schools registered all of their elementary schools. The 8 school districts with district-level DAP grants registered 20 schools for Year 8 SCAAP testing. A total of 44 DAP elementary schools from around the state participated in the Year 8 SCAAP assessments. In addition, two schools from Beaufort County School District, which received an arts-related grant other than DAP grant, participated in SCAAP Year 8 assessments. In total, approximately 4,100 students from the 46 schools participated in the Year 8 SCAAP assessments, with 3,550 of those students completed the music assessment and 3,570 of those students taking the visual arts assessment.

Training Test Administrators

Because SCAAP testing is web-based, representatives from each participating school are responsible for test administration. In Year 8, the SCAAP personnel held two in-person test administrator training sessions in February: one in Columbia at the University of South Carolina, and the other in Lexington. The Columbia session was broadcast via the Internet using a live video presentation conferencing tool known as *Breeze Live*. All new teachers participating in SCAAP were required to attend one of the two training sessions in person; the experienced teachers could choose to attend the in-person session or to get a refresher remotely via *Breeze Live*. The recording of the training session was also available on the SCAAP website (<https://scaap.ed.sc.edu/mainpages/admin.asp>). The purpose of the training sessions was to familiarize the test administrators with the SCAAP administration requirements and procedures including (a) managing registered students on the SCAAP website, (b) administering the web-based multiple-choice test forms, (c) administering performance tasks, and (d) returning the test materials. All information regarding the SCAAP assessment procedures was also documented in the *South Carolina Arts Assessment Program Test Administration Manual* from 2009. Hard copies of the manual were either provided to test administrators at the training sessions or mailed to them, and electronic copies were available on the SCAAP website. In addition, SCAAP personnel provided assistance and helped test administrators with troubleshooting during regular school hours.

Administration Procedures

Each participating school was allowed to determine the individual dates for administration of the SCAAP 4th grade music and visual arts assessments, provided that testing began no earlier than March 1st, 2009 and was completed no later than April 1st, 2009. Individual classes at each participating school were randomly assigned to take one of the two multiple-choice test forms in each arts area. All 4th grade students, or a minimum of 100 students, were required to complete the multiple-choice section of the music and visual arts assessments. Forty students from each school or 120 students from each district were required to complete the music and visual arts performance tasks.

Multiple-Choice Administration: Music & Visual Arts

Approximately one hour was allocated for completing the web-based multiple-choice section. Each music multiple-choice test form had 45 items. The music multiple-choice test forms were divided into two sections: “Understanding Music” and “Listening to Music”; the visual arts multiple-choice test forms were not divided into sections. Students worked individually on computers and were able to answer questions at their own pace. Each student wore headphones, provided by SCAAP, enabling students to listen to stimulus material and answer questions based on that material. The web-based format of the multiple-choice section allowed students to play the stimuli as many times as they wished.

Performance Task Administration: Music

The SCAAP music performance tasks were administered individually to students. The music test administrators are trained by the SCAAP personnel during the test administrator training sessions to administer the tasks following a standardized procedure. The test administrators (a) play the CD directions for each student, (b) digitally record that student’s performance using a microphone provided by the SCAAP, and (c) then save that student’s performance on a provided flash drive using a file naming convention. Test administrators return the flash drives to the SCAAP personnel using a business reply envelope provided by the SCAAP. After the files are returned, the SCAAP personnel then employ a mass uploading procedure that places the student performance task files on the SCAAP website. All procedures for administering the music performance tasks are included in the *South Carolina Arts Assessment Program Test Administration Manual* and presented in detail at the test administrator training sessions. In addition, the SCAAP personnel provided technical support during regular school hours.

On average, each student took approximately 8 minutes to complete both music performance tasks. The average time required was determined based on teacher feedback, which indicated that the time required for administering both music performance tasks ranged from 5 – 20 minutes for each student.

Performance Task Administration: Visual Arts

The SCAAP visual arts performance tasks were administered to groups of students. Each student is provided with a copy of the SCAAP Visual Arts Performance Task booklet. The visual arts test administrators are instructed by the SCAAP personnel during the test administrator training sessions to administer the tasks following a standardized procedure. Test administrators are asked to assist their students in writing their last names and identification numbers on each page of the performance task booklet and to make sure that students use only a # 2 pencil to complete their drawing. A script was provided to visual arts teachers to help them lead their students through the testing process for the performance tasks in a standardized way. For the visual arts performance tasks, a maximum of 90 minutes was allotted to complete both visual arts performance tasks—approximately 30 minutes per task (i.e., Task 1, Task 2a, and Task 2b).

After administering the visual arts performance tasks, the test administrators return the student performance tasks to the SCAAP personnel using a pre-paid, business reply envelope. Student performance task booklets are then scanned by the SCAAP personnel, saved in a .jpeg format, and uploaded to the SCAAP website via a mass uploading procedure. All procedures for administering the visual arts performance task files are included in the *South Carolina Arts Assessment Program Test Administration Manual* and presented in detail at the test administrator training sessions.

IV. SCORING

Multiple-Choice Section

For the web-based multiple-choice test forms, student responses were stored on the SCAAP website in the SQL-DB server. The SCAAP personnel organized and cleaned the SCAAP database to generate online results for the schools. The results were calculated based on the answer key stored in the item database. The SCAAP personnel also downloaded the database from the website for further statistical analysis such as classical item analysis and item response theory (IRT) analysis.

Performance Task Section

Benchmarking Music and Visual Arts Performance Tasks

Each year prior to performance task rating, Music and Visual Arts Advisory committees meet with the SCAAP personnel to review and finalize the performance task rubrics from the previous year and to benchmark student performances. In Year 8, about 100 benchmarked performances were identified for each music performance task, and 75 were identified for each visual arts performance task. Copies of the Year 8 finalized performance task rubrics are presented in Chapter 3.

Web-based Rating System and Procedure

Each year, the SCAAP personnel invite arts teachers and content area experts who have attended the Arts Assessment Institute (AAI) to serve as raters. All raters are required to attend a one-day rater training session held at a central location. The entire rating procedure, including the rater training and monitoring, is web-based via the SCAAP website. A rater manual provided technical information regarding the web-based system and the rating procedure. Very few computer problems were reported during Year 8 rating, and most raters completed the rating of all performance tasks within one month.

The student performances benchmarked during the validation sessions were divided into four sets that were used for rater training and monitoring (a) an anchor set to demonstrate proficiency at each score level for training raters, (b) a practice set to create practice tests used during rater training sessions, (c) a qualifying set to create qualifying tests and refresher tests, and (d) a seed set to be distributed among items to be rated. During the rater training session, the SCAAP personnel presented the raters with the anchor items for each task. Raters reviewed the anchor items and read the comments provided by the Validation Committee members. Then, the raters took a 10-item web-based practice test. The practice test presented immediate feedback regarding the validated scores and the Validation Committee members' comments. Finally, the raters were required to pass a 15-item randomly-generated qualifying test for each performance task. To pass, each rater needed to score at least 90% on the qualifying test before becoming eligible to rate student responses. If a rater did not pass the qualifying test after three attempts, that rater was required to discuss his/her score discrepancies with the SCAAP personnel and then retake the qualifying test until a passing score was achieved.

Subsequent to passing the qualifying tests, raters were permitted to rate student performances remotely via the website from home according to their own schedule. Based on the workload, in Year 8 each music rater was assigned to rate about 640 student responses per performance task and approximately 1,280 student responses in total; each visual arts rater was assigned about 580

student responses per performance task and approximately 1,770 student responses in total. Raters were randomly assigned to rater groups.

The SCAAP personnel monitor rater consistency throughout the remote rating process using refresher tests and seed items. Similar to the qualifying tests, refresher tests are 15-item randomly generated tests that each rater must pass with at least 90%. Raters are automatically directed to the refresher test after scoring 100 student performances or after failing three seed items. Seed items are pre-scored student performances that are distributed among un-scored student performances. A rater is considered to have failed a seed item if the score differs from the committee's score by two or more points (nonadjacent). A rater is not permitted to continue rating until he/she passes the refresher test.

The SCAAP score resolution method for nonadjacent scores utilizes expert raters, which is a widely accepted rating practice (Johnson, Penny, Fisher, & Kuhs, 2000; Johnson, Penny, Fisher, & Kuhs, 2003; Johnson, Penny, Gordon, 2001). For Year 8, raters were able to use augmentation scores (+ or -) when grading the responses. Previous studies have indicated that score augmentation tends to improve inter-rater reliability (Penny, Johnson, and Gordon, 2000). In Year 8, about one third of the visual arts performances and one third of the music performances were double-rated, meaning that each of those student performances was scored by two raters. The remaining student responses were each scored by a single rater.

V. ANALYSIS AND RESULTS

Multiple-Choice Section

Analysis

The results presented in this chapter include reliability and descriptive statistics for all music and visual arts multiple-choice test forms. Analyses were conducted using classical test theory, and computations were performed using SAS and SPSS statistical software packages. Additional analyses were also conducted using Item Response Theory (IRT) to equate the test forms based on the students' abilities. The SCAAP personnel used BILOG and IRTLRF software to perform the computations.

Reliability

All multiple-choice items were binary (i.e., scored 0 or 1). SCAAP researchers computed the classical reliability indices for each test form using Cronbach's alpha and a corrected split-half index. Furthermore, the empirical reliability based on the fitted IRT model was computed based on the variance of the ability level (θ) for both forms. The formula for the empirical reliability is $\frac{\text{Variance}(\theta)}{\text{Variance}(\theta) + \text{Variance}(\text{Error})}$. Table 5.1 contains the classical reliability indices for each test form and the empirical reliability for the equated scores.

Table 5.1

Reliability Indices for SCAAP 2008 Music and Visual Arts Multiple-Choice Items

Arts	Empirical Reliability	Test Form	Number of Items	Cronbach's Alpha	Corrected Split-Half
Music	0.84	1	45	0.82	0.82
		2	45	0.84	0.84
Visual Arts	0.87	1	45	0.86	0.86
		2	45	0.86	0.86

A minimum reliability index of .85 is necessary for a test form to be used for making high stakes decisions about individual students (Phillip, 2000). For research and evaluation purposes and for low-stake assessments, a minimum reliability index of .70 is required (Herman, Ashbacher, & Winters, 1992). According to Thorndike, Cunningham, Thorndike, & Hagen (1991), a reliability of .80 for individual scores will produce a stable mean for a group of at least 25 people.

Moreover, Hill (2002) found that increasing the group size will also increase the reliability estimates. For the SCAAP 2009 test forms, the classical reliability indices range from .82 to .86 and the empirical reliability indices for music and visual arts are .84 and .87 respectively, indicating that the results are reliable for making medium stake decisions at the school level (with more than 25 students).

Each music multiple-choice test form was divided into two sections, "Understanding Music" and "Listening to Music." The Understanding section included items with and without visual

interpretive materials; however, no aural interpretive materials were used in that section. Each question in the Listening section included aural interpretive material and many of the questions also included visual interpretive material in the form of musical notation. The reliability indices for each section are reported in Table 5.2. The Understanding Music section has a reliability index of .74 for Music Form 1 and of .76 for Music Form 2. The Listening to Music section has a reliability index of .68 for Music Form 1 and of .70 for Music Form 2. The reliability estimates for the two sections of the Music multiple-choice are not recommended for comparison, as each of the two sections has a different number of items, which can impact the reliability estimates.

Table 5.2
Reliability of “Understanding and Listening” for Music Forms 1 and 2

	Music Form 1		Music Form 2	
	Reliability	Number of Items	Reliability	Number of Items
Understanding	0.74	29	0.76	29
Listening	0.68	16	0.70	16

Each SCAAP multiple-choice item is designed to address a particular content standard based on the 2003 South Carolina Visual and Performing Arts Curriculum Standards. The SCAAP researchers computed reliability indices for each standard on each test form. Those indices are presented in Table 5.3 for each music standard and in Table 5.4 for each visual arts standard. The number of items for each standard was based on the percentages in the Table of Specifications.

Table 5.3
Reliability Indices for Music Multiple-Choice Assessment by SC Standard

	Music Form 1		Music Form 2	
	Reliability	Number of Items	Reliability	Number of Items
Standard 4	0.41	4	0.39	4
Standard 5	0.56	14	0.68	14
Standard 6	0.62	11	0.53	11
Standard 7	0.37	7	0.45	7
Standard 8	0.29	3	0.15	3
Standard 9	0.41	6	0.46	6

Table 5.4
Reliability Indices for Visual Arts Multiple-Choice Assessment by SC Standard

	Visual Arts Form 1		Visual Arts Form 2	
	Reliability	Number of Items	Reliability	Number of Items
Standard 1	0.44	7	0.50	7
Standard 2	0.54	11	0.58	11
Standard 3	0.46	7	0.59	7
Standard 4	0.46	7	0.42	7
Standard 5	0.60	8	0.49	8
Standard 6	0.54	5	0.46	5

The relatively low reliability indices can be attributed to the small number of items for each content standard. For each content standard, the number of items is determined by the table of specifications for each assessment. Although the number of items and the corresponding reliability indices for individual content standards are not sufficient to allow the reporting of multiple-choice results at the standard level, these indices have improved considerably compared to previous years' findings.

Descriptive Statistics

The mean score is 23.4 for both Music Forms 1 and 2, and for Visual Arts Form 1 and 2, the mean scores were 25.8 and 26.3, respectively. These means indicate that students correctly answered more than half of the items on each test form. For a test form of 45 items with four answer options, we estimated that students would score approximately 11 points if they guessed the answers for all items. The achieved means of 23.4-26.3 for the multiple-choice assessments indicate that the students' results are based not on chance, but on their understanding of the music and visual arts concepts covered in the assessment.

Table 5.5
Descriptive Statistics for Music and Visual Arts Test Forms

Test Forms		<i>M</i>	<i>SD</i>	<i>n</i>	Number of Items
Music Form	1	23.4	7.1	1959	45
	2	23.4	7.5	1591	45
Visual Arts Form	1	25.8	8.0	2116	45
	2	26.3	7.8	1454	45

Note: M = mean of the total test scores of all students. SD = standard deviation of the total test scores of all students. n = number of students who completed a particular test form.

IRT Model Fit

Using Item Response Theory (IRT) methods to analyze large-scale assessment data offers several advantages over classical test theory methods. The main advantage of using IRT for SCAAP is that it allows the equating of test forms based on latent variables within a single year and also across years. Another advantage is the ability to use individual item information during item review sessions.

To fully realize the benefits of IRT methods, the IRT model (i.e., one-, two-, and three-parameter logistic models) used for parameter estimation must fit the data. To examine model fit and determine the most appropriate model, the analysis of the residuals with the aid of graphs is most commonly proposed by researchers (Hambleton, Swaminathan, & Rogers, 1991, and Hambleton & Swaminathan, 1985). In the IRT analysis for Year 8 and for previous years, the SCAAP personnel used the two-parameter logistic model because it provided the best fit for the SCAAP data. A detailed description of the model fit process used by the SCAAP personnel can be found in the report *Technical Documentation for the South Carolina Arts Assessment Project (SCAAP) Year 3: Fourth Grade Music and Visual Arts Assessments* submitted to the SCDE.

Equating Test Forms

For each arts area, the SCAAP multiple-choice section consists of two parallel test forms with approximately 25 linking items between the two test forms; test forms were constructed to be parallel based on the Table of Specifications. Each test form was randomly assigned to individual classrooms within each school. Concurrent calibration method was then used to compute equated test scores for individual schools. The concurrent calibration method was chosen because this method yields more stable equated scores than other methods such as linear and equipercentile equating (Petersen, Cook, & Stocking, 1983; Hills, Subhijah, & Hirsch, 1988). This method requires the creation of a combined data set with all students and all items in the two test forms. In this combined data set, students' responses to those items not included in the test form they had taken were coded as 9 (missing). For example, if Student A completed Music Test Form 1, each item in Form 1 was coded either as 1 (correctly answered) or as 0 (incorrectly answered), while all items used in Form 2 would be 9 in the combined data set. BILOG was used as the IRT two-parameter logistic model calibration software to simultaneously estimate item parameters and ability parameters.

For Year 8, SCAAP school-level results were computed by equating across test forms and across years (SCAAP 2008 and SCAAP 2009). To identify which linking items would be used in the equated analysis, the SCAAP personnel examined the behaviors of items used in both years, including revised items. Both chi-square statistics for estimated parameter differences and the areas between Item Characteristic Curves (ICCs) were considered as criteria for detecting differential behaviors. The area between the ICCs was examined because it reflects the probability of the item scoring differently in two forms. Since all items are binary-response, this probability multiplied by the number of items represents the difference in total scores of the two forms.

First, SCAAP personnel identified those linking items with insignificant parameter differences, choosing .05 as the overall level of significance and making Bonferroni adjustments for each item. Then, SCAAP personnel examined the area between the ICCs. A cut-off value of .07 was chosen to ensure that the overall variation in total scores was no more than the Standard Error of Measurement (SEM), a measure of the expected fluctuation in scores on the same test. The SEMs of the four test forms (two each for Music and Visual Arts) were estimated to range from 2.92 to 2.98, making the cut-off approximately .07. Items with the area between ICCs of less

than .07 were considered good linking items; those with areas greater than .07 were de-linked. With a set of good linking items, individual student ability estimates were computed. Because individual test forms comprise 45 items, students' ability estimates were transformed into a 45-point scale to facilitate interpretation of results by teachers and administrators.

Technical Characteristics of Items

For every item, traditional indices such as proportion of correct responses (p -values or item difficulty values) and discrimination indices (d -values) based on point-biserial correlations were examined. In addition, differential item functioning (DIF) indices based on gender and ethnicity were computed for each item. The individual item analysis results for each item in each test form are presented in Appendices A-D.

The histograms of the p -values for each test form were constructed to investigate the distribution of p -values and are presented in Figure 5.1. The histograms indicate that most items have moderate p -values and that only a few items have very low or very high p -values.

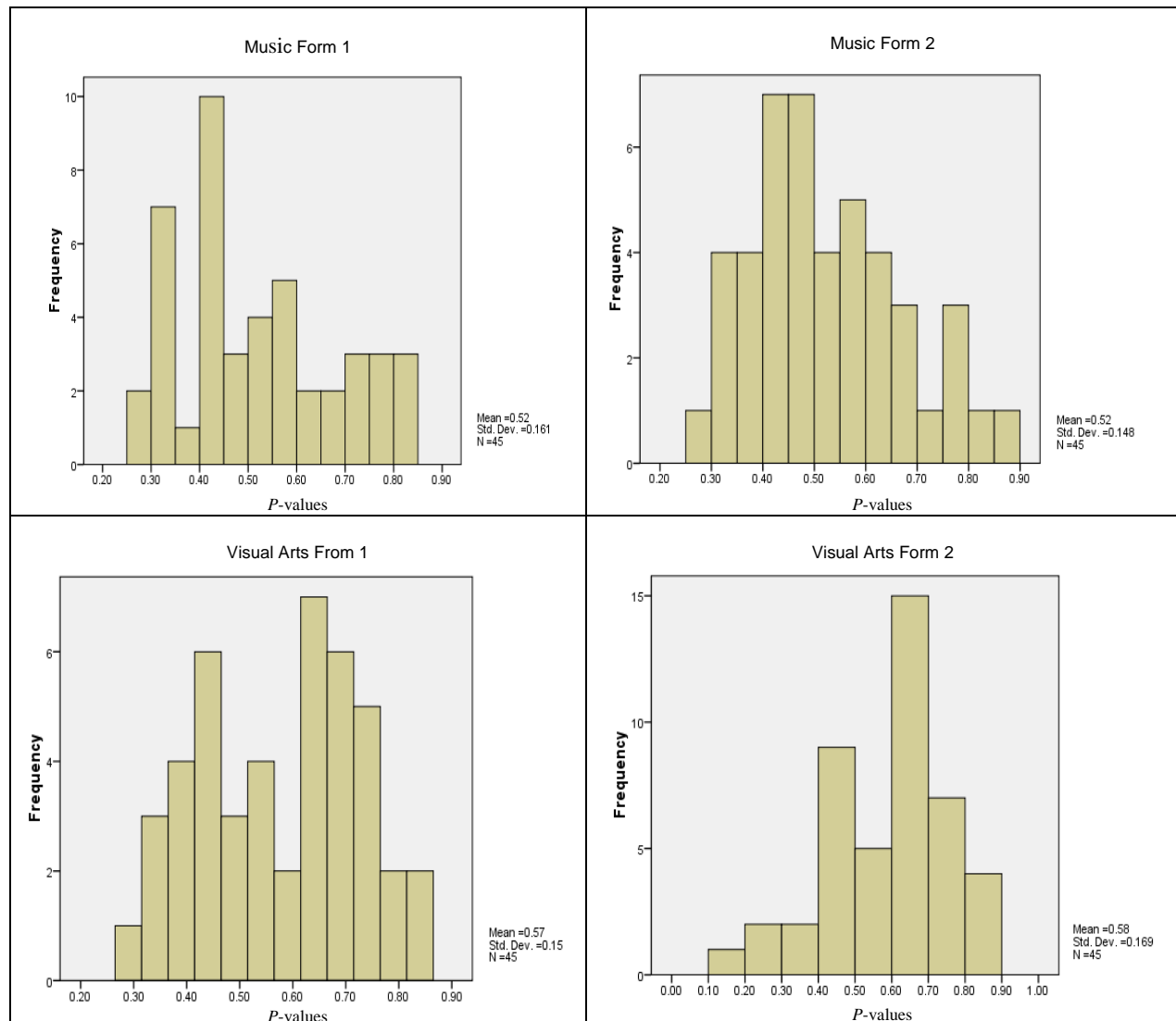


Figure 5.1. P -value Histograms for Music and Visual Arts Test Forms

The ranges of item p -values and item discrimination indices for each test form are presented in Table 5.6. Although individual item p -values range from .19 to .87, the average p -values for all test forms range from .52 to .58.

Table 5.6
Ranges of P-values and Discrimination Indices of Test Forms

Test Forms		P -value Ranges	Discrimination Indices Ranges
Music Form	1	0.28-0.83	0.11-0.49
	2	0.28-0.87	0.11-0.51
Visual Arts Form	1	0.29-0.82	0.10-0.57
	2	0.19-0.85	0.02-0.54

Individual item discrimination indices for all test items ranged from 0.02 to 0.57. Because item discrimination index is a type of correlation in which each item is correlated with the total test score, an acceptable item discrimination index means that students who answer a particular item correctly usually have an overall higher score than students who do not answer that item correctly. Conversely, a low or negative discrimination index means that students who answer a particular item correctly usually have an overall lower score than students who do not answer that item correctly. Therefore, negative or low discrimination indices are of most concern because they do not distinguish between high scoring and low scoring students, or between students who have and have not mastered the content being measured by the assessment.

According to Ebel & Frisbie (1986), items with discrimination indices (a) equal to or higher than .40 were considered to be good items, (b) between .30 and .39 were considered reasonably good items, (c) between .20 and .29 were considered marginal items that need minor revision, and (d) equal or lower than .19 were poor items that need major revision. Based on the magnitude of those individual values, each item in the SCAAP assessments was classified as having (a) an acceptable discrimination value, (b) a low discrimination value, or (c) a negative discrimination value. A summary of the number of items with questionable discrimination indices ($\leq .19$) is listed in Table 5.7. Those items with low or negative discrimination indices will be presented to the SCAAP Advisory Committee at the Item Review Session. Then, based on the committee members' recommendations, those items will either be altered or terminated in preparation for Year 9 testing.

Table 5.7
Number of Items with Low and Negative Discrimination Indices

Test Forms		Low Discrimination Indices ($0.00 \leq D \leq 0.19$)	Negative Discrimination Indices
Music Form	1	5	0
	2	3	0
Visual Arts Form	1	3	0
	2	3	0

To investigate whether the student results differ by gender or by ethnic group, the SCAAP personnel computed descriptive statistics for all test forms by gender and by ethnicity. Those descriptive statistics are presented in Table 5.8 and Table 5.9. In general, female students scored slightly higher than male students, and White students scored higher than African American students on both music and visual arts assessments. The ranges of *p*-values for each test form, which provide additional information, are presented in Table 5.10 by gender and ethnicity. Those *p*-value ranges are similar for all gender groups examined, and a slight difference exists for the ethnicity groups, with White students doing a little better than their African American counterparts.

Table 5.8
Descriptive Statistics of Test Scores by Gender

Test Forms		Gender					
		Female			Male		
		<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Music Form	1	23.6	7.0	982	23.2	7.2	977
	2	24.0	7.5	819	22.9	7.5	772
Visual Arts Form	1	26.7	7.8	1067	24.9	8.0	1049
	2	26.9	7.8	741	25.6	7.8	713

Table 5.9
Descriptive Statistics of Test Scores by Ethnicity

		Ethnicity					
		African-American			White		
Test Forms		<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Music Form	1	20.4	6.1	934	26.3	6.8	871
	2	19.7	6.3	678	26.5	7.0	796
Visual Arts Form	1	22.3	6.9	1054	29.6	7.3	913
	2	22.5	6.8	564	29.1	7.3	762

Table 5.10
Ranges of P-values of Test Forms by Gender and Ethnic Groups

Test Forms		P-value Ranges			
		Female	Male	African-American	White
Music Form	1	0.28-0.87	0.28-0.84	0.23-0.79	0.29-0.92
	2	0.26-0.87	0.28-0.86	0.25-0.78	0.28-0.94
Visual Arts Form	1	0.31-0.82	0.24-0.82	0.22-0.76	0.35-0.88
	2	0.17-0.87	0.22-0.86	0.16-0.80	0.21-0.90

To further investigate whether students' performance on the SCAAP assessments vary significantly by gender or ethnic group, all SCAAP test forms were subjected to formal Differential Item Functioning (DIF) analysis to examine each item for bias. Item bias occurs when one group with a certain ability level has an advantage over another group with the same ability level on a given item. For the SCAAP assessment, the DIF analysis involves estimating whether students in different gender or ethnic groups, matched by ability level, have the same probability of correctly responding to a particular item. If, for a particular item, either group has a different probability of correctly answering that item, the item is considered to display differential item functioning (DIF).

The DIF analysis used is based on the Mantel-Haenszel (MH) procedure, included as a routine procedure in SAS. That routine is the FREQ procedure with the Cochran-Mantel-Haenszel (CMH) option (SAS, 1990). The MH procedure is commonly used in statewide and national

standardized assessments development, such as the National Assessment of Educational Progress (NAEP) and the South Carolina statewide assessment, the Palmetto Achievement Challenge Test (PACT). Specifically, the MH procedure first categorized the students by gender or ethnicity. Then, based on their total scores, students are classified to form strata or sub-groups with approximately the same number of students in each stratum. Once the strata are formed, the proportion of students in each stratum who correctly and incorrectly answered a particular item is calculated for a “focal” group and a “reference” group. The term “focal” refers to the group of interest in a DIF analysis. For gender, the focal group refers to females and the reference group is males. For ethnicity, African-American students are the focal group and White students are the reference group.

The SAS procedure also provides an estimate of the common odds ratio and a 95% confidence interval for that ratio. Educational Testing Service (ETS) deltas were calculated by taking the natural logarithm of the common odds ratio and multiplying it with a constant, -2.35. Following are the rules developed by ETS to interpret the delta values (Zwick & Ercikan, 1989).

- “A” items are those items with an ETS delta not significantly different than zero or an absolute value of ETS delta less than 1 ($\alpha = .05$)
- “B” items are those items with an ETS delta significantly different than zero and has either (a) an absolute value of ETS delta of at least 1 but less than 1.5 or (b) at least 1 but not significantly greater than 1 ($\alpha = .05$).
- “C” items are those with ETS delta greater than 1.5 and significantly greater than 1 ($\alpha = .05$).

“A” items are considered to be free of DIF. “B” items may be used unless there are other item choices. “C” items are to be selected only if essential to meet test specifications. The direction of bias can be determined by examining the ETS delta. Only those items with negative ETS delta values are biased against the focal group.

A summary of the DIF classifications for music and visual arts items is presented in Table 5.11. All items in Year 8 were classified as either “A” or “B” items, and no item was classified as a “C” item. Individual ETS deltas for gender and ethnicity and their respective p -values for each item in each test form are presented in Appendices E-H.

Table 5.11
Summary of DIF Classification for Multiple-Choice Test Forms

Test Forms	Reference Group	Focal Group	Total N of Items	DIF Classification			
				A	B	C	
Music	1	Male	Female	45	43	2	0
		White	African-American	45	43	2	0
	2	Male	Female	45	41	4	0
		White	African-American	45	43	2	0
Visual Arts	1	Male	Female	45	45	0	0
		White	African-American	45	41	4	0
	2	Male	Female	45	41	4	0
		White	African-American	45	41	4	0
All Forms		Male	Female	180	170	10	0
		White	African-American	180	168	12	0

Performance Task Section

Students' music performance tasks were rated by three rater groups using 5-point analytic rubrics; one-third of the responses were each rated by a pair of raters and the rest were rated by two single raters. Each rater group rated approximately 1,200 performance task responses. Students' visual arts performance tasks were rated using 5-point holistic rubrics. Four raters were employed, and each of them rated approximately 1,160 performance task responses. Two raters formed one rater group pair (double-rater group) and rated about one-third of the responses; two other raters formed two single-rater groups. In the double-rater group, ratings for each student's performance task were calculated by averaging the scores of the two raters. If those two raters' scores were not adjacent, an expert rater's score was used instead. In Year 8, raters were permitted to choose augmentation scores by adding a "+" or "-" to their scores. Augmentation of scores increases or decreases a numerical score by 0.33.

Inter-Rater Reliability of Performance Tasks

Generalizability (G) theory was used to estimate inter-rater reliability for each performance task. When using G theory, a coefficient is obtained that reflects "the accuracy of generalizing from a person's observed score on a test or other measure (e.g., behavior observation, opinion survey) to the average score the person would have received under all possible conditions that the test user would be equally willing to accept" (Shavelson & Webb, 1991, p.1). The computer program

GENOVA (Crick & Brenna, 1983) was used to estimate the G-coefficients and indices of dependability to obtain information regarding sources of variation in the measurement. The index of dependability for each performance task takes into account a shift in means due to rater effects. Although multiple rater-groups were involved in scoring the performance tasks, one-facet design, P x R was used as recommended by Lee, Kantor, & Mollaun (2002). They believed that the measurement error contributed by multiple rater-groups is small as compared to the source of variation due to the examinee's ability.

Tables 5.12 and 5.13 present the inter-rater reliability estimates for the music performance tasks using the analytic rubrics and the visual arts performance tasks, respectively. Estimates are expressed as the generalizability coefficient and the index of dependability. The estimates of score reliability for each performance task were computed using a generalizability design that takes into account the variability of raters. For each performance task, the reliability estimates between 2 raters were computed.

Table 5.12

Inter-Rater Reliability of Music Performance Tasks Using Analytic Rubrics

Performance Task	Criteria	Generalizability Coefficient	Index of Dependability
1 (Singing)	Tonal	0.93	0.91
	Rhythm	0.79	0.79
	Vocal	0.84	0.81
2 (Improvisation)	Rhythm	0.90	0.88
	Improvisation	0.91	0.91

Table 5.13

Inter-Rater Reliability of Visual Arts Performance Tasks Using Holistic Rubrics

Performance Task	Generalizability Coefficient	Index of Dependability
1 (Compare and Contrast)	0.86	0.86
2a (Drawing)	0.83	0.83
2b (Self-Critique)	0.89	0.89

In Year 8, the generalizability coefficients and dependability indices are the same for all of the visual arts performance tasks, indicating no rater effect for the tasks. The generalizability coefficients and indices of dependability are exactly the same for Music Performance Task 1 Rhythm criterion and for Task 2 Improvisation criterion, but slightly different for the other three criteria with the discrepancies ranging from .02 to .03. The closeness of the generalizability

coefficients and the dependability indices suggests little to no rater effect for all SCAAP tasks. Notably, both the generalizability coefficients and dependability indices for the visual arts performance tasks increased compared to the previous year, which might be attributed to this year's successful rater training and the improvement of the visual arts rubrics.

Descriptive Statistics

The descriptive statistics for the Performance Tasks are presented in Tables 5.14 and 5.15. The task mean is simply the sum of all the examinees' scores on the task divided by the number of examinees (Johnson, Penny, & Gordon, 2008, p. 265). Similar to a multiple-choice item, the high mean of the task might indicate an easy task, and a low mean can be a sign for a difficult item.

The music analytic rubric scores indicate specific information regarding students' achieved skill level. The interpretation of scores from an analytic rubric is based on the descriptive information contained in the achieved level and in all preceding levels. See Chapter 3 for copies of the analytic rubrics used to score the SCAAP Music Performance Tasks. For Music Task 1: Singing, the mean score for the Tonal Criterion was 2.29, 3.74 for the Rhythm Criterion and 2.75 for the Vocal Criterion. Following is an interpretation summarizing the mean scores for each Task 1 criterion.

In general, students could sing the correct melodic contour, establish the tonality of the song, and almost maintain the established tonal center of the song. Most students could establish and maintain the tempo and meter of the song and perform the rhythm patterns correctly, and students could use their head voice or singing voice consistently.

For Music Task 2: Rhythm Improvisation, the mean score for the rhythm criterion was 3.20 and 2.77 for the improvisation criterion. Following is an interpretation summarizing the mean scores for each Task 2 criterion.

In general, students could establish and maintain a tempo for most of the performance, establish duple meter, and improvise an 8-beat long rhythm pattern using note values included in the prompt.

Table 5.14

Descriptive Statistics for Music Performance Tasks Using Analytic Rubrics

Task	Dimensions	<i>M</i>	<i>SD</i>	<i>N</i>
Singing	Tonal	2.29	1.40	1792
	Rhythm	3.74	0.56	1792
	Vocal	2.75	0.92	1792
	Task Total	8.78	2.41	1792
Improvisation	Rhythm	3.20	0.96	1801
	Improvisation	2.77	1.07	1801
	Task Total	5.97	1.69	1801
Total Score		14.76	3.23	1784

The descriptive statistics for the Visual Arts Performance Tasks are presented in Table 5.15. Students earned an average of 4.72 points out of 12.99 possible points for the combined visual arts performance tasks. See Chapter 3 for copies of the holistic rubrics used to score the SCAAP visual arts performance tasks. For Visual Arts Task 1 (Compare and Contrast), the mean score was 1.84 with a standard deviation of 1.16. Following is an interpretation of the mean score for Visual Arts Task 1.

In general, students demonstrated some understanding of the similarities and differences between the two artworks in the context of their writing. The average student used between two and three art terms correctly in their discussion and made mostly general or specific references to the individual artworks. Some students also explained the terms they used.

The mean score for Visual Arts Task 2a (Drawing) was 1.91 with a standard deviation of 0.84. Following is an interpretation of the mean score for Visual Arts Task 2a.

In general, students completed a drawing that addresses the assigned topic and fills the majority of the space provided. The average student made some attempt to include an environment and create depth in their drawing, but there was no distinction between the foreground and the background. The average student included little detail, texture, and pattern to enhance their drawing.

The mean score for Visual Arts Task 2b (Self-Critique) was .97 with a standard deviation of 1.23. Following is an interpretation of the mean score for Visual Arts Task 2b.

In general, students showed very limited understanding of the strengths and weaknesses of their own drawings. The average student used between one and two art terms correctly in their written critique and made specific reference to parts of their drawing when using the term(s). Some students may have included writing that was off-topic.

Table 5.15
Descriptive Statistics for Visual Arts Performance Tasks Using Holistic Rubrics

Performance Task	<i>M</i>	<i>SD</i>	<i>N</i>
1	1.84	1.16	1786
2a	1.91	0.84	1784
2b	0.97	1.23	1775
Total Score	4.72	2.43	1775

Task Difficulty Index and Discrimination Index

The task difficulty index (*p*-Value) for constructed- and extended- response items is described as “the ratio of the item mean to the item maximum possible score” (Huynh, Meyer, & Barton, 2000). The *p*-values for music performance tasks range from .68 to .70, and the *p*-values for visual arts performance tasks range from .22 to .44. The *p*-values for the music and visual arts performance tasks are presented in Table 5.16.

Table 5.16
Task Difficulty Indices for Music and Visual Arts Performance Tasks

	Music		Visual Arts		
	Task 1	Task 2	Task 1	Task 2	
				Drawing	Self-Critique
<i>p</i> -Value	0.68	0.70	0.42	0.44	0.22

Johnson, Penny, and Gordon (2008) proposed a discrimination index for performance tasks, which can be used to discriminate between low-performing examinees and high performing examinees. It is “an item-criterion correlation, with the criterion being the total raw score on both the multiple-choice and the open-ended items” (p. 271). The Pearson correlation was used to estimate the point-biserial index. The correlations for both music and visual arts performance tasks in Year 8 range from .44 to .52. A limited number of studies have been conducted to determine the standard for evaluating discrimination indices for performance tasks. As far as can be determined, only one study by Huynh, Meyer, & Barton (2000) reported that an acceptable discrimination index for performance tasks should be around .50. An acceptable discrimination index would indicate that students who do well on a particular performance task tend to score higher on the SCAAP test as a whole than students who perform relatively poorly on that task. Conversely, a low discrimination index means that students who do well on a particular performance task usually score lower than students who do not do well in that task. Based on those interpretations, the SCAAP performance tasks for Year 8 function adequately in distinguishing between high-performing and poor-performing students, with the possible exception of Music Task 2. Table 5.17 presents the discrimination indices for the tasks.

Table 5.17

Discrimination Index for Music and Visual Arts Performance Tasks

	Music		Visual Arts		
	Task 1: Singing	Task 2: Rhythm Improvisation	Task 1: Compare & Contrast	Task 2	
				Drawing	Self-Critique
Discrimination Index	0.52	0.44	0.52	0.50	0.50

Internal Validity

To investigate the internal validity of the SCAAP assessments, the SCAAP personnel computed correlations across test formats. Specifically, the SCAAP personnel calculated Pearson correlations between students' multiple-choice test scores and performance task ratings. A moderate relationship is expected among the multiple-choice tests and the performance tasks for each arts area because it is assumed that each arts assessment is measuring a similar underlying construct (i.e., music or visual arts achievement).

Table 5.18

Correlations for Music Multiple-Choice Test Forms and Performance Tasks

	Performance Task 1	Performance Task 2
Multiple-Choice Form 1	0.25	0.17
Multiple-Choice Form 2	0.25	0.30
Performance Task 1	-	0.22
Performance Task 2	-	-

The Pearson correlations between the various music assessment formats are presented in Table 5.18. For the music multiple-choice test, students' raw scores were used in the correlation computation. Analytic rubric scores were used in the correlation computation for the music performance tasks. For Music Task 1, the analytic rubric score is the sum of the three individual rubric criteria (i.e., tonal, rhythm, and vocal quality). For Music Task 2, the analytic rubric score is the sum of the two rubric criteria (i.e., rhythm and vocal quality).

The correlations between the multiple-choice section and Music Performance Task 1 are .25 for both Form 1 and Form 2. Those correlations indicate that the scores for the music multiple-choice test forms and the ratings for Music Performance Task 1 have a shared variance of approximately 6% (ρ^2). The correlations between each music multiple-choice test form and Music Performance Task 2 are .17 for Form 1 and .3 for Form 2, indicating that the scores for the music multiple-choice test forms and the scores for Music Performance Task 2 have approximately 3% to 9% variance in common. The moderately low correlations between the multiple-choice test forms and Music Performance Task indicates that the skills required to sing

a song and to improvise an 8-beat rhythm pattern are not measured using the SCAAP multiple-choice test format. Because the performance tasks are aligned with specific music standards that are not included in the multiple-choice test forms, this finding validates the inclusion of these two performance tasks in measuring standards-based music achievement.

Table 5.19
Correlations for Visual Arts Multiple-Choice Test Forms and Performance Tasks

		Performance Tasks			
		Task 1	Task 2a	Task 2b	Task 2 Sum
Multiple-Choice	Form 1	0.38	0.41	0.34	0.44
	Form 2	0.39	0.35	0.31	0.39
Performance Tasks	Task 1	-	0.28	0.37	-
	Task 2a	-	-	0.40	-

The Pearson correlations between the visual arts multiple-choice and performance task assessment formats are presented in Table 5.19. For the multiple-choice test, students' raw scores were used in the correlation computation. For the performance tasks, the holistic rubric ratings were used. The correlations between each visual arts multiple-choice test form and Visual Arts Performance Task 1 scores are .38 for Form 1 and .39 for Form 2. Those correlations indicate that the scores for the visual arts multiple-choice test forms and the ratings for Visual Arts Performance Task 1 have around 15% variance in common. The correlation between Visual Arts Form 1 and Visual Arts Performance Task 2a scores is .41, and the correlation between Form 2 and Task 2a is .35. This indicates that the Task 2a scores and the multiple-choice scores have a shared variance ranging from 12% to 16%. The correlation between Visual Arts Form 1 and Visual Arts Performance Task 2b is .34, and the correlation between Form 2 and Task 2b is .31. These results indicate that the Task 2b scores and the multiple-choice scores have a shared variance ranging from 10% to 12%. When the two portions of Visual Arts Performance Task 2 are summed and correlated with each visual arts test form, there is a correlation of .44 with multiple-choice Form 1 and a correlation of .39 with multiple-choice Form 2, with approximately 15% to 19% shared variance. The correlations between visual arts multiple-choice forms and visual arts performance tasks are moderate to moderately low, indicating that the underlying construct among test formats is similar, but that the two performance tasks provide additional information regarding students' standards-based visual arts achievement.

The Pearson correlations among the Visual Arts Performance Tasks are also presented in Table 5.19. The highest correlation among the Visual Arts Performance Tasks was observed between Task 2a and Task 2b (.40). One possible reason is that the writing for Task 2b is based on the drawing in Task 2a. Task 1 and Task 2b also demonstrated a moderate association (.37). This finding can likely be attributed to the similarities between the two tasks, which both require students to demonstrate their visual arts knowledge through writing. Moderately low correlations were found between Task 1 and Task 2a (.28).

VI. RESULTS BY SCHOOL

The reliability indices for each test form are satisfactory for providing school-level results for the SCAAP 2009 music and visual arts assessments. The following tables, Table 6.1 and Table 6.2, show the results for the schools that participated in the Year 7 (2007-2008) and Year 8 (2008-2009) testing. The multiple-choice scores provided are equated scores, which are computed using IRT. These equated scores are submitted to SCDE for reporting purposes. Individual schools receive school score reports based on assessment results for a single year. School names are not used in this report to ensure the confidentiality of participating schools.

Table 6.1

Music Assessment Results for 2008 and 2009 by School

School Code	Music Assessment 2008						Music Assessment 2009					
	MC 2008	N	Task 1	N	Task 2	N	MC 2009	N	Task 1	N	Task 2	N
S001	23.58	72	8.72	49	6.16	47	23.02	80	9.21	39	6.80	39
S002	22.58	77	-	-	-	-	-	-	-	-	-	-
S003	14.81	76	6.26	50	4.33	48	14.19	63	6.54	38	4.03	38
S004	-	-	-	-	-	-	22.33	92	9.33	40	6.32	40
S005	-	-	-	-	-	-	18.30	60	7.99	40	5.40	40
S006	17.97	78	7.57	33	5.30	36	18.71	55	9.41	39	7.11	40
S007	19.73	31	8.19	31	5.97	31	17.45	38	8.03	40	5.28	40
S008	21.72	93	7.79	52	4.48	52	24.43	109	8.62	42	4.94	42
S009	33.16	70	10.57	51	7.53	51	31.46	73	10.26	41	6.97	41
S010	30.29	112	10.74	51	7.29	49	29.98	130	9.56	41	7.08	41
S011	26.15	77	-	-	-	-	-	-	-	-	-	-
S012	-	-	-	-	-	-	15.17	39	8.51	39	5.93	39
S013	32.78	84	10.18	47	5.60	49	34.22	86	9.95	33	5.89	33
S014	28.46	74	10.04	50	6.66	50	29.30	88	10.28	40	6.40	40
S015	20.78	85	9.24	50	6.35	50	22.15	74	7.02	40	5.60	40
S016	22.22	62	8.39	58	6.05	59	25.73	51	8.66	49	5.37	50
S017	23.13	158	9.90	50	6.31	50	23.54	101	8.29	40	6.48	40
S018	28.18	80	8.39	49	7.13	50	-	-	-	-	-	-
S019	22.24	96	8.53	47	5.48	50	25.96	121	8.36	46	6.07	46
S020	27.27	101	8.70	50	6.81	50	25.25	91	9.15	40	6.60	40
S021	27.16	104	7.70	50	5.45	50	-	-	-	-	-	-
S022	27.08	83	8.21	43	5.80	43	-	-	-	-	-	-
S023	26.59	99	9.49	51	6.06	51	-	-	-	-	-	-
S024	22.92	100	8.56	51	5.56	51	26.64	97	7.49	41	5.05	41
S025	24.05	105	8.89	47	5.75	49	-	-	-	-	-	-
S026	22.00	45	8.09	45	4.80	45	-	-	-	-	-	-
S027	23.97	69	9.21	37	5.91	47	25.40	73	8.22	42	5.91	43
S028	18.57	61	7.96	47	4.31	49	19.72	70	9.42	37	6.06	39
S029	15.14	61	-	-	-	-	16.01	66	6.75	32	4.74	33
S030	26.61	67	9.50	47	7.35	50	22.30	57	9.51	40	7.44	40
S031	20.23	66	7.74	49	4.56	49	28.16	58	8.06	38	5.78	38
S032	18.25	19	7.96	20	4.74	19	20.39	24	8.21	24	4.96	24

School Code	Music Assessment 2008						Music Assessment 2009					
	MC 2008	N	Task 1	N	Task 2	N	MC 2009	N	Task 1	N	Task 2	N
S033	19.72	101	8.20	50	5.17	50	18.14	109	9.61	40	4.69	40
S034	14.63	44	8.58	44	6.22	44	-	-	-	-	-	-
S035	16.47	88	7.11	50	4.73	46	16.57	87	7.84	40	3.96	39
S036	24.91	100	7.27	47	4.82	47	-	-	-	-	-	-
S037	22.05	78	8.87	51	5.08	50	-	-	-	-	-	-
S038	22.57	96	9.17	51	5.27	51	-	-	-	-	-	-
S039	21.13	70	7.32	50	5.26	50	21.50	65	8.82	39	5.44	39
S040	24.51	65	8.25	48	6.10	50	23.27	65	9.47	40	6.32	40
S041	21.30	82	6.32	49	5.13	49	23.44	76	9.06	39	5.93	40
S042	27.61	114	8.64	51	6.07	51	-	-	-	-	-	-
S043	-	-	-	-	-	-	28.72	99	10.65	41	6.75	41
S044	-	-	-	-	-	-	25.26	97	8.14	48	6.84	47
S045	27.90	82	9.30	50	6.30	50	-	-	-	-	-	-
S046	-	-	-	-	-	-	27.59	88	7.97	43	6.53	43
S047	30.06	103	9.29	44	6.15	44	-	-	-	-	-	-
S048	19.36	131	7.01	52	5.22	54	19.52	129	7.34	42	5.61	41
S049	28.39	112	9.40	48	6.58	46	32.52	58	8.91	40	5.78	40
S050	32.56	38	8.45	37	5.51	38	-	-	-	-	-	-
S051	23.22	113	8.37	47	4.52	47	22.69	112	7.93	40	5.21	40
S052	20.06	89	10.51	43	6.39	44	20.17	98	10.20	40	7.50	40
S053	16.06	81	8.10	48	6.13	48	17.50	63	8.16	32	6.61	32
S054	18.30	72	-	-	-	-	19.41	82	8.12	41	4.77	41
S055	19.38	29	9.14	29	5.82	29	26.04	33	10.65	32	6.14	32
S056	27.42	43	9.02	43	6.25	43	27.69	37	9.02	36	5.94	36
S057	29.17	107	9.13	49	7.22	50	28.07	120	8.36	40	7.06	40
S058	29.60	104	10.01	63	6.45	62	29.79	111	10.61	41	5.83	41
S059	17.74	35	7.47	31	6.10	29	20.13	31	7.58	31	6.68	31
S060	18.25	113	10.46	50	6.78	50	18.40	105	11.02	39	6.05	42
S061	29.44	19	7.78	20	6.22	20	22.49	25	8.62	29	5.77	29
S062	18.97	67	8.79	50	6.76	50	21.30	64	8.95	38	6.46	40

Table 6.2

Visual Arts Assessment Results for 2008 and 2009 by School

School Code	Visual Arts Assessment 2008								Visual Arts Assessment 2009							
	MC 2008	N	Task 1	N	Task 2d	N	Task 2w	N	MC 2009	N	Task 1	N	Task 2d	N	Task 2w	N
S001	26.39	73	1.91	48	2.49	50	1.23	50	25.86	81	1.25	40	2.27	40	1.25	40
S002	29.56	47	2.00	50	1.76	51	0.46	50	-	-	-	-	-	-	-	-
S003	17.65	70	1.41	54	1.48	51	0.09	47	17.74	60	1.65	40	1.65	39	0.50	39
S004	-	-	-	-	-	-	-	-	29.65	90	1.66	41	1.89	41	1.02	41
S005	-	-	-	-	-	-	-	-	23.85	59	2.62	40	1.80	40	0.90	40
S006	19.98	75	0.92	52	1.41	52	0.27	52	21.31	55	1.33	41	1.39	40	0.40	39
S007	22.02	30	0.78	32	1.28	35	0.10	30	19.05	37	0.93	35	1.40	35	0.35	34
S008	23.81	90	0.96	52	1.95	53	0.31	51	25.98	102	1.92	40	2.29	40	0.94	40
S009	33.04	71	2.68	51	2.21	50	2.11	50	33.30	73	2.78	40	2.36	41	2.56	41
S010	31.99	115	2.39	51	2.42	51	0.58	51	32.88	126	2.98	40	2.38	41	1.88	41
S011	29.56	75	1.41	49	1.92	49	0.35	49	-	-	-	-	-	-	-	-
S012	-	-	-	-	-	-	-	-	19.33	48	1.82	40	1.47	40	1.35	40
S013	31.26	83	1.46	50	1.87	46	0.48	48	31.58	86	1.93	41	2.11	40	0.64	40
S014	24.13	75	2.65	50	2.15	50	1.10	50	27.18	87	2.53	40	2.16	40	1.37	40
S015	25.09	85	1.54	51	2.02	50	0.91	50	25.95	74	0.95	40	1.91	41	0.45	41
S016	23.25	63	1.41	49	1.94	46	0.63	46	26.01	51	1.60	40	1.77	40	1.01	40
S017	25.07	161	1.11	50	2.13	50	0.43	49	26.40	100	1.65	40	2.16	40	0.99	40
S018	28.32	79	1.33	49	1.95	48	0.34	47	-	-	-	-	-	-	-	-
S019	25.26	96	1.11	48	1.67	46	0.35	46	27.65	122	1.55	35	1.66	35	0.62	33
S020	28.62	101	1.89	52	2.30	51	0.99	51	28.27	94	2.37	41	2.32	41	1.50	41
S021	29.15	98	1.77	48	2.25	51	0.89	50	-	-	-	-	-	-	-	-
S022	31.74	87	1.94	51	2.14	51	0.54	52	-	-	-	-	-	-	-	-
S023	28.04	97	1.53	52	1.77	50	0.23	50	-	-	-	-	-	-	-	-
S024	25.99	114	1.30	52	1.92	52	0.73	51	28.68	100	2.28	39	2.22	40	1.25	40
S025	28.85	109	1.88	52	2.14	52	0.51	52	-	-	-	-	-	-	-	-
S026	24.70	45	1.56	45	1.68	45	0.19	44	-	-	-	-	-	-	-	-
S027	28.71	81	1.34	52	1.94	52	0.76	52	28.30	77	1.17	41	2.12	41	1.08	41
S028	20.79	60	1.63	47	1.41	45	0.54	46	22.97	70	2.48	41	1.90	41	0.73	41
S029	20.03	60	0.74	50	1.41	51	0.11	50	22.46	68	0.99	41	1.65	41	0.53	41
S030	28.90	68	3.26	51	2.25	51	0.52	51	25.11	62	2.38	40	2.25	40	1.08	40
S031	22.40	67	0.97	48	2.00	47	0.29	47	26.57	57	1.74	40	2.00	40	0.75	40
S032	23.05	20	2.42	19	1.27	19	0.36	19	21.98	48	1.08	24	1.90	21	0.76	21

School Code	Visual Arts Assessment 2008								Visual Arts Assessment 2009							
	MC 2008	N	Task 1	N	Task 2d	N	Task 2w	N	MC 2009	N	Task 1	N	Task 2d	N	Task 2w	N
S033	21.35	100	0.55	50	1.53	50	0.22	50	21.07	101	0.77	37	1.19	40	0.05	40
S034	16.45	43	1.92	44	0.92	44	0.08	30	-	-	-	-	-	-	-	-
S035	15.50	86	0.47	38	1.12	37	0.17	36	17.06	86	0.68	40	1.62	40	0.71	40
S036	27.77	100	1.86	50	2.09	50	0.54	50	-	-	-	-	-	-	-	-
S037	28.01	81	2.28	51	1.94	50	0.85	49	-	-	-	-	-	-	-	-
S038	24.68	90	1.62	51	1.74	50	0.50	50	-	-	-	-	-	-	-	-
S039	23.92	70	0.97	48	1.67	48	0.26	48	21.87	70	1.13	38	1.73	40	0.51	40
S040	28.10	65	3.04	47	2.32	47	2.25	47	26.24	66	2.85	40	2.59	40	2.48	40
S041	25.27	81	1.75	51	1.64	50	0.60	50	25.64	76	2.11	41	2.15	41	1.72	41
S042	30.25	112	1.87	50	2.59	51	1.09	51	-	-	-	-	-	-	-	-
S043	-	-	-	-	-	-	-	-	30.10	96	2.09	45	2.15	45	1.29	45
S044	-	-	-	-	-	-	-	-	27.44	97	1.54	40	1.82	40	0.63	40
S045	26.93	76	0.91	46	1.74	47	0.24	45	-	-	-	-	-	-	-	-
S046	-	-	-	-	-	-	-	-	27.59	85	1.73	41	1.94	41	0.78	41
S047	31.96	102	1.45	42	2.11	42	0.46	42	-	-	-	-	-	-	-	-
S048	22.70	123	1.20	52	1.76	50	0.34	52	22.61	136	1.39	36	1.82	36	0.79	36
S049	31.51	63	2.13	52	2.57	52	1.48	52	31.55	77	2.54	41	2.63	41	2.07	41
S050	24.50	38	1.08	38	1.87	37	0.28	37	-	-	-	-	-	-	-	-
S051	27.37	111	1.39	50	1.72	50	0.54	50	28.18	113	1.78	40	1.95	39	0.89	38
S052	23.00	93	2.65	47	2.47	47	1.01	47	19.26	97	1.89	40	1.75	40	0.67	40
S053	19.84	79	1.08	47	1.34	47	0.53	47	19.97	64	1.67	33	1.40	32	0.14	32
S054	22.33	71	0.73	49	1.46	49	0.14	49	23.20	82	2.24	40	1.96	40	0.66	38
S055	23.64	28	1.76	28	1.33	27	0.56	27	25.63	33	1.41	33	1.44	33	0.79	31
S056	26.53	43	1.64	43	1.55	43	0.39	42	27.71	37	1.78	37	1.64	37	0.82	37
S057	30.46	107	1.82	47	2.14	47	0.85	46	29.68	120	1.81	41	2.19	41	0.97	41
S058	30.41	99	2.70	50	2.38	50	0.90	49	31.17	110	2.88	40	2.16	39	1.23	39
S059	20.00	35	1.33	35	1.46	33	0.26	34	22.30	30	1.84	32	1.40	30	0.55	30
S060	20.11	109	1.68	49	1.76	49	1.01	47	21.46	103	1.91	41	1.70	41	0.87	41
S061	24.49	19	2.17	19	1.97	19	2.09	19	23.51	26	2.23	29	1.74	29	0.89	29
S062	22.50	67	1.06	50	1.84	50	0.29	50	24.83	64	1.89	41	1.43	41	0.47	41

VII. TEST ADMINISTRATOR FEEDBACK SURVEY RESULTS

Teachers who administered the 2008-2009 SCAAP Music and Visual Arts Assessments were asked to provide feedback regarding implementation logistics such as preparation for the testing, technical difficulties with testing, support provided by the SCAAP personnel, and the uses of test results. In total, 28 music teachers and 23 visual arts teachers gave comments, suggestions, and timing information for each part of the test. The following is a synthesis of the teachers' feedback beginning with general training and support, continuing with specific information regarding test administration and teachers' use of assessment results, and ending with general comments and suggestions.

Training and Support

Regarding training, 98% of respondents noted that the Test Administrator Training Session adequately prepared them to administer the SCAAP assessments. One respondent indicated feeling "a little uncertain" about administering the assessment, but felt it might be because he/she "had never officially administered a standardized music test." Of the 51 respondents, 16 attended one of the two on-site trainings, and others prepared themselves either by viewing the Webinar live from an off-campus location or online at a later date and time. Several test administrators pointed out that the online training worked well. Twenty-six of the experienced test administrators who responded to this survey reported administering the test without attending or viewing the training. Even one first-time test administrator commented that "online training would have been sufficient as long as we had access to e-mails, web-pages, etc. to answer questions."

Regarding technical support and troubleshooting, 98% of the respondents indicated that they were satisfied with the support that they received from the SCAAP personnel. Several respondents wrote positive comments, with one respondent stating that "[t]he SCAAP personnel were very helpful!" and another stating, "[w]e had issues with our own computers and some sign in issues... She was very patient with us."

Comments and Suggestions

The test administrators also provided several comments and suggestions for improving the training session. Test administrators asked that important administration information, such as test codes, be highlighted in the test administration manual or that SCAAP provide a separate sheet with a list of the most important information for test administration. In addition, Test administrators asked that the SCAAP personnel email information, like SCAAP website address, to the test administrators before the training session so they have a chance to prepare. One test administrator asked if the SCAAP personnel could invite school administrators to the training sessions so that more support could be provided from the school. Finally, one test administrator suggested making an on-site training session optional, even for first-year administrators.

Recommendations and Clarifications

Based on the teacher comments, the SCAAP personnel will make the following improvements to the training sessions and overall SCAAP support system:

1. Provide test administrators with online access to test administration manuals prior to the training.
2. Emailing test administrators with an electronic version of test administration manuals prior to the training.

3. Highlight important information in the test administration manuals and provide a separate page with Frequently Asked Questions along with the manuals.
4. Provide web-based training to all test administrators.

Although the test administrator training manual and sessions included all pertinent information regarding the testing requirements, the teacher recommendations and feedback showed that extra emphasis should be put on several issues. First, the SCAAP personnel should point out to the test administrators the page numbers for the most pertinent information such as the SCAAP website and test codes. Second, the SCAAP personnel should emphasize that each individual school is responsible for providing the appropriate accommodations for students with IEPs. Finally, although a teacher provided a valuable suggestion to include school administrators in the training session so that they could be better informed of the importance of the SCAAP tests, financial and time constraints limit the feasibility of this option. Thus, the SCAAP personnel will instead provide an informational brochure or a brief PowerPoint introduction to SCAAP upon request.

Administering the SCAAP Assessments

Of the 28 survey respondents who administered the music assessment, 71.4% also administered this assessment during the 2007-2008 academic year, 46.4% during the 2006-2007 year, 21.4% during the 2005-2006 year, and 10.7% during 2004-2005 year. Of the 23 respondents who administered the visual arts assessment, 60.9% also administered the test during the 2007-2008 academic year, 39.1% administered the test during 2006-2007, 26.1% administered this test during 2005-2006, and 8.7% administered the test during 2004-2005.

Multiple-Choice Assessment

In reference to how long it took to administer the multiple-choice assessment for music, 34.6% reported 30 to 45 minutes, 50% reported between 46 and 60 minutes, and 15.4% reported over 60 minutes. Of the respondents for visual arts, 31.8% indicated that the multiple-choice assessment took between 30 and 45 minutes, and 68.2% indicated that it took between 40 and 60 minutes to administer.

In reference to the multiple-choice assessment directions, 80% of the test administrators indicated that they did not have to provide instructions to students beyond those included in the "Test Administrator's Script" when administering the multiple-choice test. Of those who did provide additional instructions, some mentioned that they added instructions to solve computer-related problems. For example, one responded told students "don't close the window if the computer freezes" and another instructed students to "be patient and wait on the screen to load" if the website froze or loaded very slowly. Other respondents provided general test instructions that could be applied to any multiple-choice test such as "reading each entry and choices before making a choice" and "be sure that you listen to each example while looking at the written music given." Test administrators also reported telling students they could listen to the musical examples more than once.

On average, about 57.4% of the 51 respondents reported that their students experienced technical problems while taking the multiple-choice test. Of these, 11 specifically noted that at least one computer froze during testing. A few respondents indicated that the graphics were not displayed, sound files did not play, or it took a long time for the computer to load the graphics or sound files. Several respondents mentioned problems with logins, and one special case was that a student has "a twin in the same grade with a similar name...the computer said she had already taken the test." Other technical problems included that the computers kicked students off,

automatically shut down, or students mistakenly clicked “submit” button, and could not get back to the tests.

Comments and Suggestions

Only 23.7% of respondents offered suggestions for improving the implementation of the multiple-choice test. These included involving school administrators more in the test, making the test shorter, and putting the entire question including all answers on the screen at the same time so that students do not need to scroll down. Because MAP testing was arranged in many districts at the same time as SCAAP, several respondents mentioned the difficulty in scheduling a computer lab and suggested rescheduling SCAAP. One respondent pointed out that because the MAP tests present the correct answers for each question while SCAAP does not, students mistakenly thought they got all the questions correct in SCAAP. A suggestion was made to match the SCAAP test format with MAP.

Recommendations and Clarifications

Several respondents suggested that SCAAP personnel provide the special code for getting back to the middle of tests. Previously, SCAAP has not provided this code along with the actual test codes in order to prevent schools from dividing the test into several parts taken over multiple days. During normal school hours, SCAAP personnel are available to provide schools with the special code when needed. During the training sessions, test administrators are informed that they should notify the SCAAP personnel if their testing is not scheduled during normal school hours. Additional SCAAP office coverage is arranged during the requested hours. Although the availability of this special code is repeatedly discussed during the test administration training sessions, the SCAAP personnel agreed that this information should be included in the test administration manual on the FAQ section. Regarding the respondent’s suggestion that SCAAP match its test format with MAP testing, the SCAAP personnel recommend that test administrators instead remind students about the differences between the two tests prior to administering the SCAAP assessments. SCAAP is not capable of matching the MAP format because it is not a summative assessment, and it does not provide results on the individual student level.

Music Performance Task Assessment

About 50% of the test administrators reported that they used less than five minutes and about 30% reported that they used between 6 and 10 minutes to administer each Music Performance Task to each student. About 80% of the test administrators reported that the students easily understood the performance tasks directions. The majority of additional instructions involved addressing routine questions, such as instructing students when to begin performing, showing students how far they should be from the microphone, and calming students down. One test administrator specifically indicated that he/she spent more time providing additional instructions and clarification for the rhythm task because

Some [students] thought they were supposed to copy the rhythm on the CD even though the directions specified that they should create their own. I think they were feeling overwhelmed and were not listening clearly due to nerves. Also, some students did not understand that they could use ANY syllables they wanted to; they thought they could only use the syllables in the example. This is why I took more time on the rhythm portion than the singing portion.

A number of test administrators suggested providing sample materials for class practice before real testing. This suggestion will be considered by the SCAAP personnel and presented to the SCAAP Music Advisory Committee to evaluate its feasibility.

Visual Arts Performance Task Assessment

The majority (40.9%) of test administrators reported that they took less than 30 minutes to administer each Visual Arts Performance Task. None of them indicated that they spent less than 10 minutes to complete Task 1: Compare/Contrast or Task 2b: Evaluate your Drawing, but 2 respondents reported taking less than 10 minutes to complete Task 2a: Drawing. About 81% of the respondents indicated that students understood the performance task directions and no additional instructions were provided. In explaining why they needed to provide additional instructions, most test administrators reported that they wanted to be certain that students understood and followed each point in the directions. One respondent stated,

I just had to make sure they understood that there were several things being asked of them. Even though you read the steps, sometimes they just see one part of the directions. They focus on "Draw a building"; They just need to have it emphasized that when they draw whatever they draw that several things are expected to be in the picture.

In addition to reiteration and emphasis of the directions, additional instructions provided by test administrators included simple instructions about writing, not to rush through, and to ask students to use their imagination.

The survey also asked test administrators their suggestions for improving the administration procedures for the performance tasks. The suggestions included the following: (1) provide more time for completing the tasks, or leave the time limit open-ended, like the PASS test, (2) provide a rubric about what was expected from students. To address these suggestions, the SCAAP personnel will reiterate during the test administrator training in the future that there is no time limit for the completion of performance tasks, and that a majority of students in the past have finished each task within 30 minutes. Additionally, the rubrics for rating the Visual Arts Performance Tasks were provided to test administrators along with the Visual Arts Test Administrator Script for the first time this year and will continue to be provided to test administrators in the future.

This was the first year that the Visual Arts Performance Tasks were administered using a standardized script. Most respondents thought the script was helpful for administering the Visual Arts performance tasks, which "gave the setting a more structured feel like the PASS test". One of the test administrators stated, "(t)he script was well-written and explained the directions thoroughly. It was a wonderful asset to the test." However, one respondent thought the script might not be necessary, because "the test application was simple enough."

Using the SCAAP Assessment Results

Test administrators also explained how they planned to use the results from the SCAAP assessment. Overall, respondents indicated that the assessments were useful for locating the weaknesses and strengths in their teaching, and they felt this identification of strengths and weaknesses would improve individual instruction and guide curriculum planning in the following year. The majority of respondents noted that they use the assessments to improve classroom instruction. They stated that the tests show students' strengths and weaknesses and

help teachers pinpoint standards that need more attention in their classroom in the future. In addition, a respondent noted he/she would “try to bring in more outside arts related resources that students are not normally exposed to.” One respondent mentioned that the assessments are used for comparisons with other schools, and one respondent mentioned that the tests are used for grants. Another respondent noted that the assessments themselves, but not necessarily the official results, could also be used “as an audition for the school chorus for the following year.”

Only three respondents provided critical responses related to the SCAAP assessment results. One respondent wrote, “(t)each to the test I guess! ... to be uniform in what we teach across the state?” Another respondent complained about the content of the tests,

I only see the students for 45 minutes each week. There is no way that I can cover all of the possible questions you could ask. I spend lots of time reviewing. I cannot cover every artist, every culture, every art style, etc., sufficiently. If we question a test item, you pat response is that it is covered by your standards. I may cover all of the standards, but I may not cover them with the same information you did. To answer your question, I just cover as much as I can in the short amount of time that I have. Since there is no specific reference as to the question or questions that were missed the most, I don't know a specific area to review.

The third respondent replied with the following critique:

I was disappointed in last year's results. They were very vague and broad. It would be helpful to see the individual student scores. I also am not sure how each standard is being addressed on the test, therefore how I am to teach appropriate material? Shouldn't children be assessed in the way in which they learn the material? So, to answer the question, I'm not sure how I plan to use the results because they aren't specific enough to give us any helpful feedback.

Those concerns indicated to the SCAAP personnel that additional explanation regarding the development of the SCAAP assessments is needed to correct arts educators’ misunderstandings regarding the interpretation of the assessment results. These comments indicate that respondents did not understand that SCAAP is not designed to inform teachers about individual students’ performance. Consequently, the SCAAP personnel recommend providing additional assessment development training for arts educators.

Additional Comments/Suggestions

Finally, the survey asked test administrators for any additional comments or suggestions regarding the SCAAP assessments. Several respondents mentioned experiencing scheduling difficulties due to time conflicts with PASS testing and MAP testing and limited availability of the computer lab. Two teachers expressed concern that the assessments involve too much reading and writing ability, which might not truly reflect their students’ achievements in music or visual arts. Several test administrators pointed out the importance of getting individual scores, scores from other schools, and changes in SCAAP results over time to better inform planning future curriculum and instruction. For the online test, one teacher suggested changing the color for the option chosen, because

where it shows what questions have been answered, they showed up in blue. Some other test the students had must have had a similar thing, where all the questions

answered correctly showed up in blue. They all said the test was too easy, because they got all the answers right!

One respondent thanked SCAAP staff for “not having to upload the scores as time for an arts teacher is more and more constricted,” and another thanked SCAAP for providing such a test that “I am very proud for the arts to be tested in such a format so that we are held to the same standards as regular education teachers.”

Test administrators suggested several future changes to the assessment, though many of these suggestions have already been addressed in relation to previous survey questions. Those suggestions were as follows: (1) change the scheduling to after/before PASS or MAP, (2) provide individual results, (3) provide results from other participating schools, (4) provide explanation of school rank in SCAAP, (5) notify school administrators about importance of SCAAP tests, (6) change the way the option that was chosen showed up on the screen, (7) provide several practice tests, (8) improve readability in the test and change the font for some questions.

VIII. REFERENCES

- Breland, H. (1983). The direct assessment of writing skill: A measurement review. *Technical Report No. 83-6*. Princeton, NJ: College Entrance Examination Board.
- Crick, G. E., & Brennan, R. L. (1983). *Manual for GENOVA: A generalized analysis of variance system (ACT Tech. Bulletin No. 43)*. Iowa City, IA: American College Testing Program.
- Ebel, R.L., & Frisbie, D.A. (1986). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Gordon, E. E. (2002). *Rating scales and their uses for measuring and evaluating achievement in music performances*. GIA Publications, Inc.
- Herman, J., Ashbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hills, J.R., Subhiyah, R.G., & Hirsch, H.J. (1991). Equating minimum-competency tests: Comparisons of methods. *Journal of Educational Measurement* 25, 221-231.
- Hill, R. (2002, April). Examining the reliability of accountability systems. Presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Huynh, H., Meyer, P., & Barton, K. (2000). *Technical documentation for the 1999 Palmetto Achievement Challenge Tests of English language arts and mathematics, grades three through eight*. Columbia: University of South Carolina.
- Johnson, R.L., Penny, J., Fisher, S., & Kuhs, T. (2000). The relationship between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education* 13(2), 121-138.
- Johnson, R.L., Penny, J., Fisher, S., & Kuhs, T. (2003). Score resolution: Investigation of the reliability and validity of resolved scores. *Applied Measurement in Education* 16(4), 299-322.
- Johnson, R.L., Penny, J., Gordon, B. (2001). Score resolution and the interrater reliability of holistic scores in rating essays. *Written Communication* 18(2), 229-249.
- Johnson, R., Penny, J., & Gordon, B. (2008). *Assessing performance: Developing, scoring, and validating performance tasks*. New York: Guilford Publications.
- Lee, Y., Kantor, R., & Mollaun, P. (April, 2001). *Score dependability of the writing and speaking sections of new TOEFL*. Paper presented as a part of the symposium titled "Research in Support of the Development of New TOEFL" at the annual meeting of National Council on Measurement in Education (NCME) in New Orleans, LA.

- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Penny, J., Johnson, R.L., Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education*, 68(3), 269-288.
- Petersen, N.S., Cook, L.L., & Stocking, M.L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8, 137-156.
- Phillips, S. E. (2000). "Legal Corner: GI Forum v TEA." *NCME Newsletter* 8, no.2 (April): n.p.
- SAS Institute (1990). *SAS User's Guide*. Cary, NC: Author
- SCDE (1999). *South Carolina Visual and Performing Arts Standards*. SC: South Carolina Department of Education.
- SCDE (2002). *South Carolina Visual and Performing Arts Curriculum Standards 2002 Field Review Draft*. SC: South Carolina Department of Education.
- SCDE (2003). *South Carolina Visual and Performing Arts Curriculum Standards 2003*. SC: South Carolina Department of Education.
- Shavelson, R.J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- Thorndike, R., Cunningham, G., Thorndike, R. & Hagen, E. (1991). *Measurement and evaluation in psychology and education*, 5th edition. New York: MacMillan Publishing Company.
- Zwick, R. & Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement*, 26(1), 55-66.

IX. APPENDICES

APPENDIX A: 2009 SCAAP Item Analysis Results for Music Form 1

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			A	B	C	D		
1	0.83	0.35	82*	4	8	4	B	A
2	0.62	0.43	11	11	62*	14	A	A
3	0.78	0.42	11	77*	6	4	A	A
4	0.73	0.35	11	72*	6	9	A	A
5	0.43	0.37	16	10	29	42*	A	A
6	0.35	0.26	14	35*	29	19	A	A
7	0.49	0.31	49*	7	19	23	A	A
8	0.76	0.38	2	16	4	75*	A	A
9	0.43	0.27	13	43*	37	5	B	A
10	0.53	0.46	29	3	53*	13	A	A
11	0.80	0.35	6	3	80*	9	A	A
12	0.56	0.32	8	56*	12	23	A	A
13	0.31	0.13	30*	7	34	27	A	A
14	0.51	0.42	50*	13	16	19	A	A
15	0.33	0.27	34	33*	14	17	A	A
16	0.30	0.16	12	41	30*	14	A	A
17	0.65	0.41	10	64*	13	11	A	A
18	0.42	0.30	19	42*	19	18	A	A
19	0.43	0.29	18	17	20	43*	A	B
20	0.51	0.39	9	18	51*	19	A	A
21	0.31	0.34	23	37	8	30*	A	A
22	0.71	0.40	2	71*	2	23	A	A
23	0.42	0.34	42*	20	21	16	A	A
24	0.68	0.49	68*	12	11	7	A	A
25	0.28	0.13	28*	11	41	18	A	A
26	0.53	0.28	22	53*	14	9	A	A
27	0.38	0.30	17	37*	22	22	A	A
28	0.34	0.15	24	34*	26	14	A	A
29	0.31	0.36	23	23	22	31*	A	A
30	0.44	0.35	44*	9	26	19	A	A
31	0.72	0.37	4	4	73*	17	A	A
32	0.47	0.38	47*	24	6	21	A	A
33	0.60	0.41	12	10	16	60*	A	A
34	0.57	0.25	8	21	13	57*	A	A
35	0.42	0.33	16	10	42*	30	A	A
36	0.46	0.24	47*	18	9	25	A	A
37	0.40	0.32	23	40*	15	20	A	A
38	0.61	0.44	62*	11	14	11	A	A
39	0.59	0.43	21	7	11	60*	A	A
40	0.32	0.11	20	15	31	32*	A	A
41	0.42	0.39	22	42*	16	18	A	A
42	0.57	0.37	57*	16	10	15	A	A
43	0.43	0.38	43*	26	11	18	A	A
44	0.83	0.44	4	85*	5	4	A	B
45	0.75	0.37	11	76*	5	5	A	A

APPENDIX B: 2009 SCAAP Item Analysis Results for Music Form 2

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			A	B	C	D		
1	0.64	0.41	11	11	63*	13	A	A
2	0.79	0.38	10	78*	2	7	B	A
3	0.59	0.28	6	10	23	59*	A	A
4	0.48	0.33	12	12	26	47*	A	A
5	0.69	0.28	14	68*	12	4	A	A
6	0.40	0.25	12	39*	29	17	A	A
7	0.77	0.35	3	14	4	77*	A	A
8	0.44	0.40	34	5	44*	15	A	A
9	0.42	0.35	13	41*	39	5	B	A
10	0.65	0.36	17	16	1	64*	B	A
11	0.50	0.32	49*	8	18	23	A	A
12	0.37	0.39	36*	25	20	16	A	A
13	0.42	0.51	41*	17	9	31	A	B
14	0.44	0.27	44*	27	19	8	A	A
15	0.56	0.36	10	55*	13	19	A	A
16	0.36	0.38	33	36*	14	15	A	A
17	0.69	0.35	8	68*	12	10	A	A
18	0.46	0.39	18	46*	15	19	A	A
19	0.50	0.39	15	16	17	50*	A	A
20	0.32	0.19	14	41	31*	12	A	A
21	0.54	0.41	10	15	53*	19	A	A
22	0.28	0.11	15	28*	32	23	A	A
23	0.58	0.44	21	57*	13	7	A	A
24	0.38	0.34	12	26	22	37*	A	A
25	0.59	0.36	19	59*	12	9	A	A
26	0.62	0.39	62*	14	7	14	A	A
27	0.43	0.33	21	22	43*	11	A	A
28	0.32	0.14	32*	15	35	16	A	A
29	0.35	0.36	20	24	20	34*	B	A
30	0.46	0.34	45*	8	25	20	A	A
31	0.73	0.39	4	5	73*	17	A	A
32	0.47	0.36	47*	20	5	26	A	A
33	0.37	0.31	24	37*	16	22	A	A
34	0.53	0.29	9	22	14	53*	A	A
35	0.40	0.46	28	20	10	40*	A	A
36	0.45	0.38	15	45*	12	26	A	A
37	0.46	0.37	20	46*	15	17	A	A
38	0.75	0.49	9	8	6	75*	A	A
39	0.61	0.41	22	5	10	62*	A	A
40	0.32	0.14	19	17	29	32*	A	A
41	0.46	0.40	47*	23	10	18	A	A
42	0.58	0.39	22	58*	11	8	A	A
43	0.64	0.36	10	8	64*	17	A	A
44	0.87	0.41	4	88*	3	3	A	B
45	0.80	0.39	9	81*	4	5	A	A

APPENDIX C: 2009 SCAAP Item Analysis Results for Visual Arts Form 1

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			A	B	C	D		
1	0.82	0.36	1	12	3	82*	A	A
2	0.63	0.47	5	17	13	63*	A	B
3	0.73	0.42	11	9	72*	6	A	A
4	0.67	0.38	10	67*	8	13	A	A
5	0.78	0.45	77*	12	4	5	A	A
6	0.73	0.39	5	73*	12	8	A	A
7	0.82	0.25	6	7	81*	4	A	A
8	0.70	0.41	69*	15	10	4	A	A
9	0.76	0.34	10	75*	2	11	A	A
10	0.72	0.10	7	4	72*	15	A	B
11	0.62	0.38	8	15	14	61*	A	A
12	0.68	0.34	12	11	68*	8	A	A
13	0.71	0.42	19	4	5	71*	A	A
14	0.60	0.37	12	59*	15	12	A	A
15	0.34	0.16	12	33*	17	35	A	A
16	0.66	0.46	4	22	66*	6	A	B
17	0.37	0.39	20	25	16	37*	A	A
18	0.48	0.50	11	48*	11	28	A	A
19	0.75	0.30	7	7	75*	9	A	A
20	0.63	0.49	4	62*	7	26	A	A
21	0.51	0.21	22	19	5	51*	A	A
22	0.65	0.42	7	7	65*	19	A	A
23	0.44	0.36	31	43*	11	13	A	A
24	0.44	0.41	26	43*	16	12	A	A
25	0.64	0.51	8	9	17	64*	A	A
26	0.62	0.40	14	12	10	62*	A	A
27	0.61	0.41	9	17	60*	12	A	A
28	0.45	0.43	32	45*	15	6	A	A
29	0.77	0.43	6	5	77*	11	A	A
30	0.37	0.41	26	37*	17	18	A	A
31	0.56	0.32	10	10	56*	23	A	A
32	0.54	0.57	7	26	12	53*	A	A
33	0.43	0.30	18	17	20	43*	A	A
34	0.69	0.42	68*	8	16	6	A	A
35	0.45	0.34	45*	20	16	17	A	A
36	0.29	0.22	24	22	28*	24	A	A
37	0.69	0.29	69*	15	11	2	A	A
38	0.56	0.45	15	9	19	56*	A	A
39	0.41	0.45	40*	31	14	13	A	B
40	0.47	0.43	18	47*	19	15	A	A
41	0.42	0.36	19	41*	17	21	A	A
42	0.37	0.11	11	35	36*	15	A	A
43	0.34	0.31	19	27	18	34*	A	A
44	0.33	0.36	32*	17	35	14	A	A
45	0.55	0.50	20	13	10	54*	A	A

APPENDIX D: 2009 SCAAP Item Analysis Results for Visual Arts Form 2

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			A	B	C	D		
1	0.85	0.26	5	4	4	85*	A	A
2	0.67	0.43	5	16	11	67*	A	A
3	0.76	0.39	5	16	1	75*	A	B
4	0.70	0.34	8	69*	7	14	A	A
5	0.76	0.40	7	76*	7	8	A	A
6	0.76	0.30	6	75*	7	10	A	A
7	0.84	0.35	3	3	84*	9	B	A
8	0.73	0.38	73*	15	7	3	A	A
9	0.58	0.42	57*	23	12	6	A	A
10	0.68	0.26	10	12	9	67*	A	A
11	0.66	0.40	7	14	11	66*	A	A
12	0.69	0.30	11	11	69*	7	A	A
13	0.76	0.43	16	3	4	75*	A	A
14	0.64	0.37	11	64*	12	11	A	A
15	0.41	0.33	11	41*	24	22	A	A
16	0.72	0.45	4	18	72*	4	A	A
17	0.41	0.41	21	20	16	40*	A	A
18	0.52	0.54	8	52*	11	27	A	A
19	0.80	0.38	2	4	80*	11	A	A
20	0.65	0.53	3	65*	6	23	A	A
21	0.51	0.39	22	1	51*	24	A	A
22	0.74	0.44	13	74*	6	4	A	A
23	0.46	0.41	45*	13	13	27	A	B
24	0.45	0.38	29	44*	14	10	A	A
25	0.69	0.49	6	8	16	69*	B	B
26	0.63	0.45	16	10	9	63*	A	A
27	0.61	0.46	10	16	60*	12	A	A
28	0.61	0.46	16	10	12	60*	A	A
29	0.80	0.39	5	4	80*	9	A	A
30	0.40	0.40	25	39*	17	17	A	A
31	0.56	0.38	10	8	56*	24	A	B
32	0.30	0.29	30*	26	18	24	A	A
33	0.19	0.16	19*	42	15	22	B	A
34	0.70	0.44	70*	8	16	4	A	A
35	0.52	0.42	27	9	11	52*	B	A
36	0.44	0.34	23	44*	17	14	A	A
37	0.70	0.28	70*	11	13	3	A	A
38	0.63	0.48	12	8	15	63*	A	A
39	0.32	0.17	36	19	11	31*	A	A
40	0.47	0.39	18	47*	19	14	A	A
41	0.41	0.38	18	41*	14	25	A	A
42	0.23	0.02	25	23*	21	29	A	A
43	0.40	0.22	15	26	17	40*	A	A
44	0.32	0.36	32*	16	35	15	A	A
45	0.60	0.50	18	10	10	32*	A	A

APPENDIX E: 2009 SCAAP Item P-Values for Music Test Forms by Gender

Item	<u>Music Form 1</u>			<u>Music Form 2</u>		
	Female	Male	ETS Δ	Female	Male	ETS Δ
1	0.87	0.79	1.30	0.66	0.61	0.23
2	0.66	0.58	0.93	0.84	0.73	1.30
3	0.81	0.74	0.95	0.63	0.56	0.58
4	0.76	0.69	0.90	0.49	0.46	0.13
5	0.42	0.43	-0.23	0.68	0.69	-0.30
6	0.38	0.33	0.43	0.43	0.37	0.39
7	0.52	0.47	0.46	0.78	0.76	-0.17
8	0.78	0.74	0.48	0.43	0.46	-0.70
9	0.37	0.49	-1.33	0.38	0.46	-1.05
10	0.51	0.55	-0.65	0.61	0.69	-1.25
11	0.80	0.81	-0.39	0.52	0.47	0.32
12	0.56	0.56	-0.09	0.38	0.35	0.00
13	0.32	0.29	0.26	0.40	0.43	-0.73
14	0.47	0.54	-0.93	0.44	0.44	-0.16
15	0.35	0.32	0.23	0.58	0.53	0.35
16	0.31	0.29	0.19	0.38	0.34	0.12
17	0.68	0.62	0.59	0.72	0.65	0.47
18	0.42	0.42	-0.09	0.49	0.44	0.31
19	0.45	0.41	0.33	0.53	0.48	0.21
20	0.51	0.52	-0.16	0.31	0.33	-0.39
21	0.29	0.32	-0.56	0.52	0.56	-0.85
22	0.70	0.72	-0.38	0.26	0.30	-0.63
23	0.45	0.39	0.54	0.57	0.58	-0.54
24	0.70	0.67	0.21	0.37	0.38	-0.41
25	0.28	0.29	-0.19	0.57	0.61	-0.70
26	0.54	0.52	0.11	0.66	0.58	0.67
27	0.38	0.37	0.04	0.45	0.41	0.16
28	0.35	0.34	0.11	0.33	0.31	0.22
29	0.34	0.28	0.63	0.40	0.28	1.17
30	0.45	0.43	0.07	0.47	0.44	0.04
31	0.74	0.71	0.18	0.77	0.68	0.91
32	0.46	0.48	-0.39	0.47	0.47	-0.31
33	0.58	0.63	-0.74	0.41	0.33	0.68
34	0.57	0.56	-0.03	0.54	0.52	0.04
35	0.42	0.42	-0.18	0.40	0.41	-0.49
36	0.47	0.46	-0.01	0.45	0.44	-0.19
37	0.38	0.42	-0.55	0.45	0.47	-0.49
38	0.60	0.63	-0.50	0.78	0.72	0.50
39	0.58	0.61	-0.39	0.62	0.61	-0.23
40	0.33	0.31	0.30	0.34	0.31	0.28
41	0.41	0.44	-0.39	0.51	0.42	0.73
42	0.55	0.58	-0.39	0.57	0.59	-0.59
43	0.45	0.42	0.22	0.66	0.61	0.29
44	0.83	0.84	-0.50	0.87	0.86	-0.20
45	0.77	0.73	0.48	0.84	0.76	0.89

Note: The item number indicates the position of the item in the test form and not the item content.

APPENDIX F: 2009 SCAAP Item P-Values for Music Test Forms by Ethnicity

Item	Music Form 1			Music Form 2		
	African-American	White	ETS Δ	African-American	White	ETS Δ
1	0.79	0.88	0.38	0.55	0.71	0.44
2	0.52	0.72	-0.46	0.72	0.85	0.25
3	0.70	0.87	-0.58	0.54	0.65	0.25
4	0.67	0.77	0.27	0.42	0.55	0.38
5	0.37	0.49	0.29	0.60	0.74	-0.21
6	0.34	0.38	0.81	0.37	0.42	0.67
7	0.44	0.56	0.17	0.70	0.83	0.27
8	0.69	0.83	-0.32	0.37	0.51	0.53
9	0.39	0.48	0.30	0.35	0.48	0.23
10	0.42	0.64	-0.28	0.53	0.74	-0.70
11	0.74	0.87	-0.30	0.42	0.56	0.13
12	0.50	0.61	0.34	0.27	0.45	-0.29
13	0.28	0.32	0.28	0.25	0.57	-1.35
14	0.43	0.58	0.30	0.37	0.49	-0.09
15	0.28	0.40	-0.17	0.46	0.63	-0.03
16	0.28	0.32	0.21	0.25	0.46	-0.59
17	0.55	0.74	-0.38	0.60	0.75	0.00
18	0.35	0.49	-0.19	0.36	0.55	-0.16
19	0.41	0.45	1.21	0.44	0.57	0.61
20	0.40	0.62	-0.62	0.28	0.33	0.29
21	0.23	0.37	-0.32	0.41	0.64	-0.58
22	0.64	0.77	0.22	0.31	0.28	0.89
23	0.35	0.50	-0.08	0.46	0.67	-0.14
24	0.57	0.80	-0.82	0.29	0.43	0.00
25	0.27	0.29	0.27	0.52	0.64	0.44
26	0.48	0.58	0.03	0.50	0.72	-0.62
27	0.31	0.44	-0.04	0.35	0.50	0.05
28	0.33	0.36	0.38	0.31	0.32	0.61
29	0.24	0.38	0.01	0.29	0.39	0.93
30	0.38	0.50	0.35	0.40	0.50	0.68
31	0.65	0.80	-0.31	0.61	0.82	-0.75
32	0.37	0.58	-0.50	0.37	0.55	-0.25
33	0.49	0.71	-0.59	0.31	0.43	0.27
34	0.52	0.61	0.17	0.46	0.59	0.05
35	0.38	0.45	0.80	0.28	0.50	-0.27
36	0.44	0.49	0.64	0.35	0.52	-0.09
37	0.34	0.46	0.07	0.35	0.55	-0.53
38	0.54	0.70	0.18	0.63	0.84	-0.31
39	0.47	0.71	-0.77	0.48	0.72	-0.86
40	0.30	0.33	0.15	0.29	0.35	-0.14
41	0.31	0.54	-0.77	0.37	0.53	0.24
42	0.48	0.65	-0.22	0.50	0.64	0.46
43	0.36	0.51	0.13	0.52	0.71	-0.45
44	0.75	0.92	-1.31	0.78	0.94	-1.05
45	0.68	0.82	-0.22	0.72	0.86	0.11

Note: The item number indicates the position of the item in the test form and not the item content.

APPENDIX G: 2009 SCAAP Item P-Values for Visual Arts Test Forms by Gender

Item	Visual Arts Form 1			Visual Arts Form 2		
	Female	Male	ETS Δ	Female	Male	ETS Δ
1	0.82	0.82	-0.62	0.87	0.83	0.37
2	0.65	0.62	-0.27	0.68	0.66	-0.10
3	0.76	0.69	0.42	0.77	0.75	-0.20
4	0.70	0.65	0.15	0.71	0.69	-0.17
5	0.80	0.75	0.16	0.79	0.73	0.41
6	0.78	0.68	0.79	0.78	0.73	0.29
7	0.82	0.81	-0.24	0.82	0.86	-1.42
8	0.73	0.66	0.49	0.76	0.70	0.32
9	0.79	0.72	0.51	0.58	0.57	-0.29
10	0.75	0.69	0.50	0.71	0.65	0.40
11	0.65	0.59	0.22	0.70	0.63	0.50
12	0.71	0.65	0.18	0.71	0.67	0.24
13	0.72	0.70	-0.23	0.77	0.74	-0.12
14	0.60	0.59	-0.31	0.65	0.64	-0.17
15	0.31	0.36	-0.78	0.39	0.43	-0.77
16	0.70	0.62	0.46	0.76	0.67	0.87
17	0.39	0.36	-0.13	0.40	0.41	-0.45
18	0.50	0.47	-0.23	0.52	0.52	-0.50
19	0.78	0.72	0.36	0.82	0.79	0.22
20	0.63	0.62	-0.59	0.67	0.63	0.02
21	0.51	0.51	-0.28	0.52	0.50	-0.08
22	0.69	0.61	0.45	0.76	0.73	-0.08
23	0.44	0.43	-0.31	0.50	0.42	0.52
24	0.43	0.44	-0.56	0.46	0.43	-0.08
25	0.69	0.59	0.61	0.76	0.62	1.52
26	0.66	0.58	0.55	0.67	0.59	0.49
27	0.64	0.58	0.20	0.64	0.57	0.36
28	0.48	0.43	0.09	0.66	0.55	0.81
29	0.79	0.75	0.02	0.83	0.78	0.40
30	0.37	0.37	-0.55	0.41	0.38	0.00
31	0.59	0.54	0.20	0.57	0.55	-0.13
32	0.55	0.52	-0.49	0.30	0.30	-0.32
33	0.41	0.46	-0.80	0.17	0.22	-1.03
34	0.70	0.67	-0.17	0.72	0.69	0.01
35	0.49	0.42	0.39	0.49	0.55	-1.12
36	0.34	0.24	0.87	0.44	0.45	-0.50
37	0.69	0.69	-0.37	0.70	0.71	-0.57
38	0.59	0.53	0.12	0.67	0.58	0.67
39	0.43	0.39	-0.12	0.31	0.32	-0.26
40	0.50	0.45	0.08	0.50	0.44	0.35
41	0.44	0.40	-0.02	0.44	0.39	0.20
42	0.40	0.34	0.42	0.22	0.25	-0.51
43	0.34	0.35	-0.44	0.40	0.40	-0.31
44	0.33	0.32	-0.33	0.33	0.32	-0.36
45	0.56	0.53	-0.21	0.65	0.55	0.77

Note: The item number indicates the position of the item in the test form and not the item content.

APPENDIX H: 2009 SCAAP Item P-Values for Visual Arts Forms by Ethnicity

Item	Visual Arts Form 1			Visual Arts Form 2		
	African-American	White	ETS Δ	African-American	White	ETS Δ
1	0.76	0.88	0.07	0.80	0.88	-0.29
2	0.57	0.70	1.07	0.59	0.73	0.46
3	0.64	0.83	-0.43	0.65	0.84	-1.03
4	0.59	0.76	-0.28	0.62	0.75	0.08
5	0.68	0.88	-0.80	0.67	0.83	-0.45
6	0.66	0.81	-0.04	0.68	0.80	-0.40
7	0.76	0.87	-0.45	0.77	0.90	-0.69
8	0.63	0.77	0.49	0.65	0.79	-0.18
9	0.69	0.82	0.18	0.46	0.66	-0.24
10	0.74	0.71	1.01	0.63	0.72	0.27
11	0.54	0.70	0.17	0.59	0.72	0.33
12	0.63	0.73	0.55	0.63	0.73	0.18
13	0.61	0.82	-0.65	0.69	0.81	0.58
14	0.52	0.66	0.36	0.56	0.69	0.24
15	0.31	0.37	0.07	0.32	0.48	-0.31
16	0.54	0.80	-1.03	0.62	0.80	-0.10
17	0.29	0.45	0.38	0.34	0.46	0.75
18	0.34	0.64	-0.92	0.36	0.64	-0.61
19	0.69	0.82	-0.12	0.77	0.85	0.87
20	0.51	0.74	-0.15	0.53	0.73	0.31
21	0.46	0.57	-0.15	0.43	0.59	0.16
22	0.57	0.76	-0.19	0.65	0.82	-0.45
23	0.38	0.49	0.88	0.40	0.50	1.18
24	0.32	0.56	-0.69	0.30	0.54	-0.93
25	0.55	0.75	0.39	0.60	0.74	1.05
26	0.56	0.69	0.59	0.55	0.70	0.45
27	0.52	0.71	-0.15	0.51	0.69	0.06
28	0.33	0.57	-0.55	0.51	0.67	0.25
29	0.70	0.85	0.13	0.74	0.85	0.25
30	0.30	0.46	0.59	0.30	0.48	0.05
31	0.47	0.67	-0.62	0.42	0.68	-1.26
32	0.41	0.68	-0.20	0.22	0.36	-0.32
33	0.35	0.53	-0.37	0.16	0.21	0.24
34	0.61	0.77	0.40	0.61	0.76	0.47
35	0.39	0.51	0.54	0.39	0.62	-0.71
36	0.22	0.35	-0.56	0.34	0.52	-0.34
37	0.61	0.78	-0.72	0.62	0.76	-0.44
38	0.47	0.66	0.28	0.53	0.70	0.42
39	0.33	0.49	1.01	0.27	0.35	-0.20
40	0.38	0.58	0.13	0.36	0.55	-0.32
41	0.32	0.52	-0.27	0.31	0.48	-0.12
42	0.35	0.40	-0.05	0.26	0.22	0.68
43	0.29	0.40	0.36	0.37	0.40	0.73
44	0.23	0.43	-0.42	0.24	0.39	-0.10
45	0.43	0.67	-0.15	0.47	0.70	-0.25

Note: The item number indicates the position of the item in the test form and not the item content.