South Carolina
Department of Education
Together, we can.

TECHNICAL DOCUMENTATION FOR THE

SOUTH CAROLINA ARTS ASSESSMENT PROJECT (SCAAP):

ENTRY-LEVEL MUSIC & VISUAL ARTS ASSESSMENT 2008

A report submitted to

R. Scot Hockman
Education Associate for the Visual and Performing Arts
South Carolina State Department of Education

Prepared by:

Ching Ching Yap, Research Assistant Professor

Min Zhu, Research Assistant

Ashlee Lewis, Research Assistant

Tara Pearsall, Research Assistant

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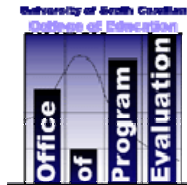College of Education
University of South Carolina

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

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. 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 tests include four entry level and two intermediate level 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

The first year of the project (Year 1) 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. Also in Year 2, the SCAAP personnel 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.

Also in Year 3, Advisory Committees were formed 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 beginning 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.

Currently, in Year 7, SCAAP comprises 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. All SCAAP assessments are administered during the spring semester, and school-level results are 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. Ching Ching Yap, perform tasks related to the educational measurement aspects of the program as well as the administrative aspects of the program. Dr. Yap 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

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 multi-media 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 sfw (video files). Many of 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 decided on 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 (http://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.

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, along 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.

9

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 7.

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	28	29
Listening to Music	17	16
Total	45	45

Music Performance Task Section: Format & Scoring
In Year 1, the Music Advisory Committee and SCAAP personnel develop two music performance tasks, which were the same tasks used in Year 7. 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 two trained raters using hierarchical analytic rubrics. 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[1] 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 7 music performance tasks are presented in Tables 2.3 to 2.7.

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

In 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. 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 same 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. Each test administrator is responsible for making sure students use a # 2 pencil and work independently.

Since Year 1, the SCAAP visual arts performance tasks have been scored by two 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 the Year 7 visual arts performance tasks are presented in Tables 2.10 to 2.12.

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

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

Table 2.11
SCAAP Visual Arts Task 2a Rubric—Drawing

Rating	The writing 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	Most of the student's writing demonstrates a clear understanding of the strengths and weaknesses of the drawing. 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, and at least 2 of the 4 terms are explanations. Only one additional term may be used incorrectly.
3	Most of the student's writing demonstrates a clear understanding of the strengths and weaknesses of the drawing. 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, and at least 1 of the 4 terms is an explanation.
2	Some of the student's writing demonstrates a clear understanding of the strengths and weaknesses of the drawing. 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, and the terms used may or may not include explanations.
1	Student's writing demonstrates limited understanding of the strengths and weaknesses of the drawing. 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, and the terms used may or may not include explanations.
0	Student uses at most 1 term correctly and clearly shows in the context of the writing that he/she understands the art term used. Most writing is off topic, or there is an insufficient amount of writing to evaluate the student's visual arts knowledge.

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 7, 28 schools and 10 school districts received Distinguished Arts Program (DAP) grants from the South Carolina State Department of Education (SCDE). All 28 schools that received school-level DAP grants were required to participate in the SCAAP 4th grade music and visual arts assessments. Each of the 10 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 registered all their elementary schools. The 10 school districts with district-level DAP grants registered 26 schools for Year 7 SCAAP testing. A total of 54 DAP elementary schools from around the state participated in the Year 7 SCAAP assessments. In addition, two schools from Beaufort County School District, that received an arts-related grant other than DAP grant, participated in SCAAP Year 7 assessments. In total, approximately 4,700 students 56 schools participated in the Year 7 SCAAP assessments, with 4,481 of those students taking the music assessment and 4,386 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 7, the SCAAP personnel held six test administrator training sessions across South Carolina during the months of January and February. The purpose of the training sessions was to familiarize the test administrators with the SCAAP administration requirements and procedures including (a) registering 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 2008. Hard copies of the manual were provided to test administrators at the training sessions 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, 2008 and was completed no later than April 4th, 2008. 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. Fifty students from each school or 150 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 in were divided into two sections: "Understanding Music" and "Listening to Music"; the visual arts multiple-choice test forms had only one section. 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 necessary.

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 flashdrive using a file naming convention. Test administrators return the flash drives to the SCAAP personnel using 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 required approximately 8 minutes to complete both music performance tasks. The average time required is based on the results from a teacher feedback survey that 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 to 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. 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 on 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 finalize the performance task rubrics from the previous year and to benchmark student performances. In Year 7, about 100 benchmarked performances were identified for each music and visual arts performance task. Copies of the Year 7 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 7 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 7 each music rater was assigned to rate about 600 student responses per performance task and approximately 1,200 student responses in total; each visual arts rater was assigned about 500

student responses per performance task and approximately 1,500 student responses in total. Raters were randomly assigned to rater groups which scored the same set of student responses.

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 (non-adjacent). 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, a widely accepted rating practice (Johnson, Penny, Fisher, & Kuhs, 2000; Johnson, Penny, Fisher, & Kuhs, 2003; Johnson, Penny, Gordon, 2001). For Year 7, raters could 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 7, about 75% of the performances in each arts area were rated by two raters.

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 researchers used BILOG and IRTLRDIF 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 $\frac{Variance(\theta)}{Variance(\theta) + Variance(Error)}$. 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	0.85	1	45	0.83	0.84
		2	45	0.83	0.84
Visual Arts	0.85	1	45	0.83	0.83
		2	45	0.84	0.84

A minimum reliability index of .85 is necessary if a test form to be used for making high stakes decisions about individual students (Phillip, 2000). For research purposes and 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 2008 test forms, the classical reliability indices range from .83 to .84 and the empirical reliability indices for music and visual arts are both .85 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 .78 for Music Form 1 and of .76 for Music Form 2. The Listening to Music section has a reliability index of .65 for Music Form 1 and of .71 for Music Form 2.

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	0.78	28	0.76	29
Listening	0.65	17	0.71	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	0.34	4	0.30	4
Standard 5	0.58	13	0.66	14
Standard 6	0.66	12	0.69	11
Standard 7	0.41	7	0.38	7
Standard 8	0.32	3	0.31	2
Standard 9	0.45	6	0.43	7

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	0.41	7	0.47	7
Standard 2	0.41	11	0.38	11
Standard 3	0.58	8	0.60	8
Standard 4	0.47	7	0.40	7
Standard 5	0.52	7	0.57	7
Standard 6	0.40	5	0.40	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 scores for Music Forms 1 and 2 are 24.3 and 24.5, respectively, and for Visual Arts Form 1 and 2 were 24.4 and 24.6. 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 24.3-24.6 points for multiple-choice tests indicate that the students' results are based on their understanding of the music and visual arts concepts and based on chance.

Table 5.5
Descriptive Statistics for Music and Visual Arts Test Forms

Test Forms		*M*	*SD*	*n*	Number of Items
Music Form	1	24.3	7.2	2404	45
	2	24.5	7.4	2077	45
Visual Arts Form	1	24.4	7.1	2472	45
	2	24.6	7.4	1914	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 assessments 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 for 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 has to 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 previous years' IRT analysis and in Year 7 IRT analysis, 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 combine data set with all students and all items in the two test forms. In this data set, students' responses to those items not included in the test form they had taken were coded as 9. For example, if Student A completed Music Test Form 1, all the items in Form 1 were coded either as 1 (correctly answered) or as 0 (incorrectly answered), while all the items used in Form 2 only would show as 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 7, SCAAP school-level results were computed by equating across test forms and across years (SCAAP 2007 and SCAAP 2008). 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 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 the items are all binary-response, this probability multiplied by the number of items would then represent 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 and then 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.
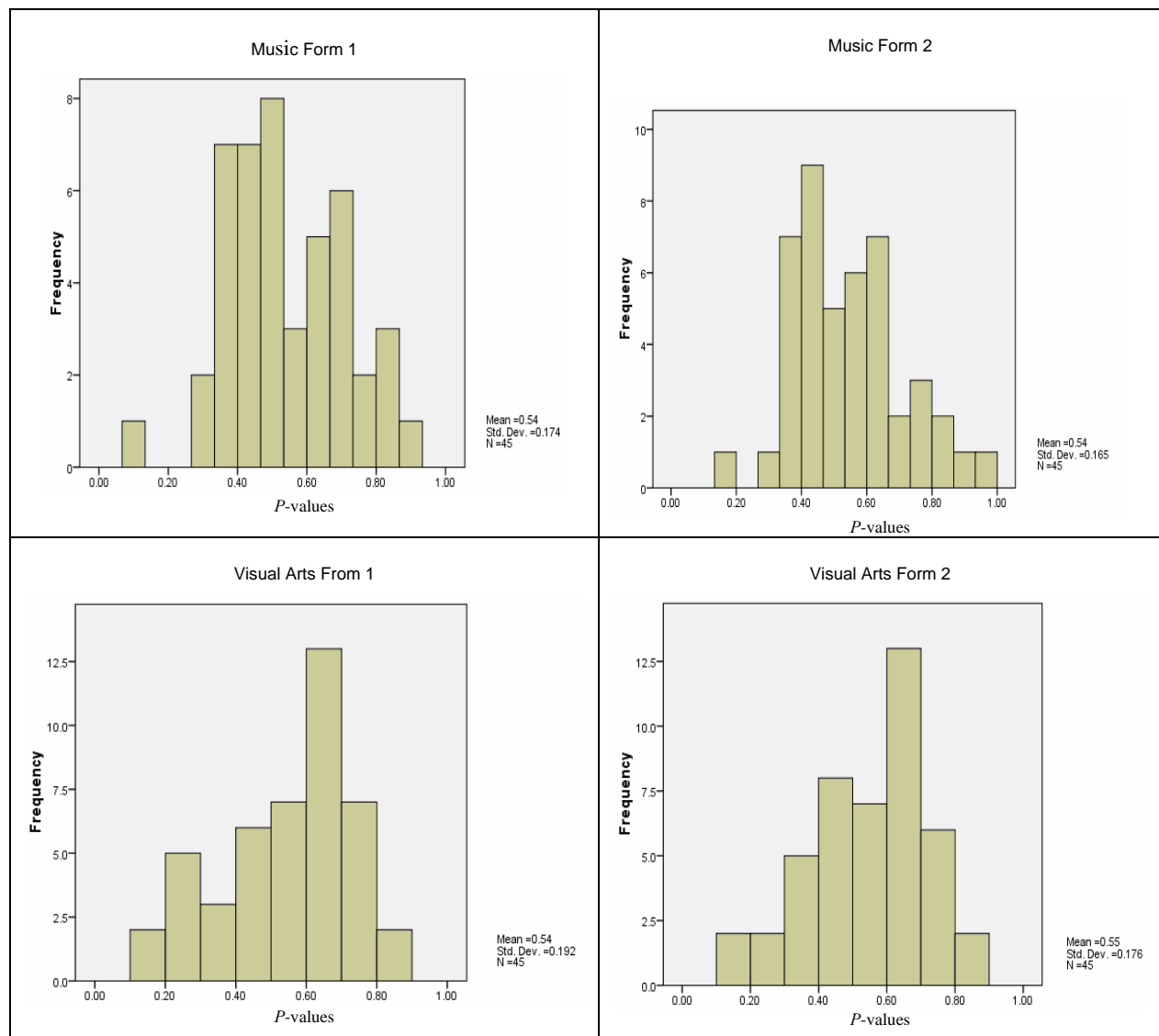
Figure 5.1. P-value Histograms for Music and Visual Arts Test Forms

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 .13 to .94, the average *p*-values for all test forms range from .54 to .55.

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

Test Forms		*P*-value Ranges	Discrimination Indices Ranges
Music Form	1	0.13-0.92	0.07-0.49
	2	0.18-0.94	0.15-0.51
Visual Arts Form	1	0.15-0.82	0.00-0.51
	2	0.16-0.84	-0.02-0.57

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 8 testing.

Table 5.7

Number of Items with Low and Negative Discrimination Indices

Test Forms		Low Discrimination Indices (0.00 ≤ D ≤ 0.19)	Negative Discrimination Indices
Music Form	1	1	0
	2	2	0
Visual Arts Form	1	7	0
	2	5	1

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 the gender groups examined, and a slight difference is detected for the ethnicity groups.

Table 5.8

Descriptive Statistics of Test Scores by Gender

		Gender					
		Female			Male		
Test Forms		*M*	*SD*	*n*	*M*	*SD*	*n*
Music Form	1	24.9	7.2	1195	23.6	7.2	1209
	2	24.9	7.4	1021	24.0	7.3	1056
Visual Arts Form	1	25.2	6.9	1218	23.6	7.1	1254
	2	25.3	7.2	959	23.9	7.5	955

Table 5.9
Descriptive Statistics of Test Scores by Ethnicity

Test Forms		African-American			White		
		M	*SD*	*n*	*M*	*SD*	*n*
Music Form	1	21.1	6.5	956	26.7	6.7	1268
	2	21.2	6.5	896	27.4	6.9	1044
Visual Arts Form	1	21.1	6.4	968	27.0	6.4	1300
	2	21.2	6.5	821	27.5	6.8	981

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	0.13-0.94	0.13-0.91	0.13-0.89	0.13-0.95
	2	0.18-0.96	0.18-0.92	0.17-0.91	0.18-0.97
Visual Arts Form	1	0.15-0.83	0.15-0.82	0.12-0.76	0.16-0.88
	2	0.16-0.85	0.16-0.83	0.16-0.80	0.17-0.89

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 that same ability level for a particular 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, each 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 by 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 (α = .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 (α = .05).
- Items "C" items are those with ETS delta greater than 1.5 and significantly greater than 1 (α =.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. Most items were classified as either "A" or "B" items and only 1 item in Music Form 1 was classified as a "C" item. Typically, "C" items are reviewed for possible sources of bias. This particular item, however, has been used on the SCAAP assessments since 2004 and was classified as an "A" item in all previous years. For this year's administration, a minor change was made to the options in that item to avoid repeating a word (see Item-Writing Guidelines in Chapter 2). Because the content of that item has not been modified, the change in the DIF classification from "A" to "C" may be due to chance and not potential bias. That item will be reviewed by the Advisory Committee in Year 8. Individual ETS deltas for gender and ethnicity and its 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	44	1	0
		White	African-American	45	41	3	1
	2	Male	Female	45	42	3	0
		White	African-American	45	43	2	0
Visual Arts	1	Male	Female	45	44	1	0
		White	African-American	45	41	4	0
	2	Male	Female	45	43	2	0
		White	African-American	45	45	0	0
All Forms		Male	Female	180	173	7	0
		White	African-American	180	170	9	1

Performance Task Section

Students' music performance tasks were rated by four rater groups using 5-point analytic rubrics; 75% of the responses were each rated by a pair of raters and the rest were rated by a single rater. Each rater group rated approximately 1,200 performance tasks. Students' visual arts performance tasks were rated using 5-point holistic rubrics. Nine raters were employed, and each of them rated approximately 1,000 performance tasks. Eight raters formed four rater group pairs and one rater formed a single-rater group. In the double-rater groups, ratings for each student's performance task were calculated by averaging the scores of the two raters in each pair. If those two raters' scores were not adjacent, an expert rater's score was used instead. In Year 7, 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-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	0.93
	Rhythm	0.72	0.72
	Vocal	0.84	0.83
2 (Improvisation)	Rhythm	0.90	0.90
	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	0.74	0.74
2a (Drawing)	0.78	0.76
2b (Self-Critique)	0.85	0.85

The generalizibility coefficients and dependability indices are the same for each music performance task criteria, with the exception of the vocal criteria that had a .01 discrepancy between the generalizability coefficient and the index of dependability. The generalizability coefficients and indices of dependability are the exactly the same for Visual Arts Task 1 and Task 2b, but slightly different for Visual Arts Task 2a. The closeness of the generalizibility coefficients and the dependability indices suggests little to no rater effect for all SCAAP tasks.

Notably, both the generalizibility coefficients and dependability indices are lower for the visual arts performance tasks than for music performance tasks. This result is consistent with previous research indicating that inter-rater reliability is typically lower in holistic scoring than in analytic scoring (Breland, 1983).

Descriptive Statistics
The descriptive statistics for the Music Performance Tasks are presented in Table 5.14. 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 2.31, 3.69 for the rhythm criterion, and 2.66 for the vocal criterion. Following is an interpretation summarizing the mean scores for each Task 2 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 3.18 and 2.69 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	*M*	*SD*	*N*
Singing	Tonal	2.31	1.43	2413
	Rhythm	3.69	0.69	2413
	Vocal	2.66	0.95	2413
	Task Total	8.66	2.66	2413
Improvisation	Rhythm	3.18	1.00	2417
	Improvisation	2.69	1.11	2417
	Task Total	5.86	1.71	2417
Total Score		14.53	3.50	2379

The descriptive statistics for the Visual Arts Performance Tasks are presented in Table 5.15. Students earned an average of 4.13 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, the mean score was 1.61 with a standard deviation of 1.17. 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 1.88 with a standard deviation of 0.85. 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 0.62 with a standard deviation of 0.98. 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	M	SD	N
1	1.61	1.17	2629
2a	1.88	0.85	2611
2b	0.62	0.98	2577
Total Score	4.13	2.23	2553

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 .72 to .73, and the *p*-values for Visual Arts Performance Tasks range from .15 to .47. 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 2	
	Task 1	Task 2	Task 1	Drawing	Self-Critique
p-Value	0.72	0.73	0.40	0.47	0.15

Johnson, Penny, and Gordon (2008) summarized 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 7 range from .44 to .58. Limited 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 means that students who do well in a particular performance task usually score higher on the SCAAP test as a whole than students who perform relatively poorly in that task. Conversely, a low discrimination index means that students who get a high score in 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 7 function adequately in distinguishing between high-performing and poor-performing students, with the possible exception of Visual Arts Task 2b. 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 2	
	Task 1: Singing	Task 2: Rhythm Improvisation	Task 1: Compare & Contrast	Drawing	Self-Critique
Discrimination Index	0.58	0.47	0.54	0.49	0.44

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	0.32	0.21
Multiple-Choice Form 2	0.28	0.29
Performance Task 1	-	0.28
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 .32 and .28 for Form 1 and Form 2, respectively. 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 8% to 10% (ρ^2). The correlations between each music multiple-choice test form and Music Performance Task 2 are .21 for Form 1 and .29 Form 2, indicating that the scores for the music multiple-choice test forms and the scores for Music Performance Task 2 have approximately 4% to 8% variance in common. The moderately low correlation between the multiple-choice test forms and Music Performance Task 2 indicates that the skill required to improvise an 8-beat rhythm pattern is not measured using the SCAAP multiple-choice test format. Furthermore, 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	0.40	0.36	0.31	0.40
	Form 2	0.36	0.39	0.25	0.39
Performance Tasks	Task 1	-	0.26	0.39	0.40
	Task 2a	-	-	0.32	-

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 .40 for Form 1 and .36 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 approximately 13% to 16% variance in common. The correlation between Visual Arts Form 1 and Visual Arts Performance Task 2a scores is .36, and the correlation between Form 2 and Task 2a is .39. This indicates that the Task 2a scores and the multiple-choice scores have a shared variance ranging from 13% to 15%. The correlation between Visual Arts Form 1 and Visual Arts Performance Task 2b is .31, and the correlation between Form 2 and Task 2b is .25. These results indicate that the Task 2b scores and the multiple-choice scores have a shared variance ranging from 6 to 10%. When both portions of Visual Arts Performance Task 2 are summed and correlated with each visual arts test form, there is a correlation of .40 with multiple-choice Form 1 and a correlation of .39 with multiple-choice Form 2, with approximately 15% to 16% 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 1 and Task 2b (.39). This finding can likely be attributed to the similarities between the two tasks, which both require students to demonstrate visual arts knowledge through writing. Moderately low correlations were found between Task 1 and Task 2a (.26) and between Task 2a and Task 2b (.32).

VI. RESULTS BY SCHOOL

The reliability indices for each test form are satisfactory for providing school level results for the SCAAP 2008 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 6 (2006-2007) and Year 7 (2007-2008) 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 confidentiality of participating schools.

Table 6.1

Music Assessment Results for 2007 and 2008 by School

School Code	Music Assessment 2007						Music Assessment 2008					
	MC 2007	N	Task 1	N	Task 2	N	MC 2008	N	Task 1	N	Task 2	N
S001	-	-	-	-	-	-	24.47	43	9.02	43	6.25	43
S002	12.37	72	6.40	40	4.15	40	13.05	76	6.26	50	4.33	48
S003	29.18	86	10.30	48	5.15	46	28.93	84	10.18	47	5.60	49
S004	26.84	66	8.84	50	5.60	50	25.28	112	9.40	48	6.58	46
S005	16.29	131	6.35	51	5.03	51	17.24	131	7.01	52	5.22	54
S006	-	-	-	-	-	-	23.49	99	9.49	51	6.06	51
S007	20.43	81	7.43	48	5.11	49	22.25	100	7.27	47	4.82	47
S008	23.70	100	7.41	47	6.09	47	-	-	-	-	-	-
S009	22.46	44	7.62	41	5.48	44	28.67	38	8.45	37	5.51	38
S010	22.90	78	8.04	48	6.38	48	21.44	69	9.21	37	5.91	47
S011	23.47	44	7.70	41	6.60	42	-	-	-	-	-	-
S012	25.27	45	9.23	44	6.52	44	-	-	-	-	-	-
S013	18.54	83	7.30	46	5.43	51	-	-	-	-	-	-
S014	17.30	20	8.19	16	4.92	18	16.38	19	7.96	20	4.74	19
S015	16.46	24	7.75	18	6.30	23	15.98	35	7.47	31	6.10	29
S016	20.91	80	7.45	43	5.00	45	19.54	78	8.87	51	5.08	50
S017	22.87	46	7.37	46	4.91	46	-	-	-	-	-	-
S018	18.02	90	7.64	36	6.31	36	-	-	-	-	-	-
S019	22.53	123	9.89	49	5.82	50	20.70	158	9.90	50	6.31	50
S020	24.74	74	8.70	48	6.41	50	25.29	74	10.04	50	6.66	50
S021	21.65	51	6.58	50	5.09	51	-	-	-	-	-	-
S022	17.03	78	7.66	48	5.59	48	-	-	-	-	-	-
S023	23.14	54	7.96	52	6.36	54	23.13	77	-	-	-	-
S024	24.96	60	9.96	50	7.32	50	-	-	-	-	-	-
S025	17.96	100	6.80	49	4.31	49	-	-	-	-	-	-
S026	-	-	-	-	-	-	-	-	-	-	-	-
S027	-	-	-	-	-	-	20.62	100	8.56	51	5.56	51
S028	21.27	105	-	-	-	-	20.91	72	8.72	49	6.16	47
S029	19.19	87	7.03	47	4.70	47	-	-	-	-	-	-
S030	-	-	-	-	-	-	18.28	85	9.24	50	6.35	50
S031	-	-	-	-	-	-	18.53	70	7.32	50	5.26	50
S032	14.85	78	-	-	3.51	47	14.70	88	7.11	50	4.73	46

41

School Code	Music Assessment 2007						Music Assessment 2008					
	MC 2007	N	Task 1	N	Task 2	N	MC 2008	N	Task 1	N	Task 2	N
S033	12.98	105	6.62	49	5.34	50	-	-	-	-	-	-
S034	14.11	43	7.70	41	5.02	41	12.85	44	8.58	44	6.22	44
S035	-	-	-	-	-	-	24.20	104	7.70	50	5.45	50
S036	12.62	29	6.00	21	4.79	24	-	-	-	-	-	-
S037	-	-	-	-	-	-	24.68	114	8.64	51	6.07	51
S038	14.51	78	6.23	48	4.56	48	17.63	101	8.20	50	5.17	50
S039	23.92	91	7.40	48	4.58	48	-	-	-	-	-	-
S040	20.77	63	7.87	50	5.42	49	-	-	-	-	-	-
S041	22.55	48	7.96	48	6.29	48	21.97	65	8.25	48	6.10	50
S042	19.28	125	7.03	50	4.18	48	-	-	-	-	-	-
S043	18.55	101	6.69	49	4.60	50	-	-	-	-	-	-
S044	20.85	81	7.34	49	4.65	47	19.10	93	7.79	52	4.48	52
S045	28.55	74	9.71	50	7.38	50	29.35	70	10.57	51	7.53	51
S046	14.22	63	-	-	-	-	-	-	-	-	-	-
S047	25.34	80	8.53	47	6.15	46	-	-	-	-	-	-
S048	-	-	-	-	-	-	17.92	66	7.74	49	4.56	49
S049	-	-	-	-	-	-	19.56	45	8.09	45	4.80	45
S050	17.52	89	6.61	47	5.98	48	16.40	113	10.46	50	6.78	50
S051	25.03	63	9.91	48	6.05	38	-	-	-	-	-	-
S052	17.11	70	6.98	49	5.48	50	13.28	61	-	-	-	-
S053	-	-	-	-	-	-	26.90	103	9.29	44	6.15	44
S054	22.44	111	8.33	50	5.95	44	-	-	-	-	-	-
S055	24.52	68	9.24	50	6.45	50	-	-	-	-	-	-
S056	17.55	35	7.89	35	6.07	35	-	-	-	-	-	-
S057	20.33	26	7.20	25	4.65	24	17.23	29	9.14	29	5.82	29
S058	23.41	53	9.24	35	7.07	34	23.54	67	9.50	47	7.35	50
S059	16.98	83	6.26	50	4.04	49	-	-	-	-	-	-
S060	15.99	93	9.08	48	5.46	39	17.90	89	10.51	43	6.39	44
S061	17.59	43	7.45	40	5.29	42	19.71	62	8.39	58	6.05	59
S062	15.39	120	5.59	50	4.86	50	-	-	-	-	-	-
S063	20.68	72	6.09	49	4.37	51	18.60	82	6.32	49	5.13	49
S064	18.88	50	7.17	44	5.43	44	-	-	-	-	-	-
S065	15.07	68	6.78	49	4.85	49	-	-	-	-	-	-
S066	27.31	102	9.58	50	4.89	50	-	-	-	-	-	-

42

School Code	Music Assessment 2007						Music Assessment 2008					
	MC 2007	N	Task 1	N	Task 2	N	MC 2008	N	Task 1	N	Task 2	N
S067	-	-	-	-	-	-	24.35	101	8.70	50	6.81	50
S068	-	-	-	-	-	-	20.20	77	-	-	-	-
S069	13.26	76	7.26	50	5.25	50	-	-	-	-	-	-
S070	17.94	92	7.00	47	5.22	50	-	-	-	-	-	-
S071	16.70	46	8.64	45	4.97	45	16.66	61	7.96	47	4.31	49
S072	18.00	74	6.42	50	6.42	49	17.17	67	8.79	50	6.76	50
S073	-	-	-	-	-	-	24.07	83	8.21	43	5.80	43
S074	20.79	105	8.03	46	5.02	48	-	-	-	-	-	-
S075	-	-	-	-	-	-	26.58	104	10.01	63	6.45	62
S076	21.52	96	9.45	51	6.18	51	19.96	96	9.17	51	5.27	51
S077	22.28	98	8.09	49	5.54	50	-	-	-	-	-	-
S078	16.98	107	7.32	52	7.15	52	-	-	-	-	-	-
S079	21.26	35	7.62	34	4.40	35	-	-	-	-	-	-
S080	-	-	-	-	-	-	24.92	82	9.30	50	6.30	50
S081	19.00	39	7.86	37	3.77	35	-	-	-	-	-	-
S082	27.96	105	10.30	50	6.23	50	27.01	112	10.74	51	7.29	49
S083	23.96	105	6.96	50	6.74	50	25.79	107	9.13	49	7.22	50
S084	16.74	39	5.36	37	3.97	37	-	-	-	-	-	-
S085	22.79	95	10.14	50	6.72	50	-	-	-	-	-	-
S086	13.56	91	7.34	49	6.82	49	14.33	81	8.10	48	6.13	48
S087	-	-	-	-	-	-	19.75	96	8.53	47	5.48	50
S088	24.80	65	9.49	47	6.67	49	-	-	-	-	-	-
S089	-	-	-	-	-	-	16.34	78	7.57	33	5.30	36
S090	15.36	26	7.88	25	5.69	26	25.88	19	7.78	20	6.22	20
S091	20.31	80	7.85	30	5.50	30	-	-	-	-	-	-
S092	24.07	96	8.60	50	6.86	50	25.16	80	8.39	49	7.13	50
S093	17.99	46	8.59	45	6.67	45	-	-	-	-	-	-
S094	25.26	51	7.69	49	6.57	49	-	-	-	-	-	-
S095	15.21	40	6.21	39	4.72	39	-	-	-	-	-	-
S096	17.26	39	5.84	37	6.39	38	17.64	31	8.19	31	5.97	31
S097	16.02	79	6.29	49	5.04	49	16.40	72	-	-	-	-
S098	21.83	66	9.53	45	6.81	49	-	-	-	-	-	-
S099	-	-	-	-	-	-	21.46	105	8.89	47	5.75	49
S100	-	-	-	-	-	-	20.73	113	8.37	47	4.52	47

Table 6.2

Visual Arts Assessment Results for 2007 and 2008 by School

School Code	Visual Arts Assessment 2007								Visual Arts Assessment 2008							
	MC 2007	N	Task 1	N	Task 2d	N	Task 2w	N	MC 2008	N	Task 1	N	Task 2d	N	Task 2w	N
S001	-	-	-	-	-	-	-	-	26.59	43	1.64	43	1.55	43	0.39	42
S002	14.25	70	0.53	49	1.46	49	0.09	49	16.89	70	1.41	54	1.48	51	0.09	47
S003	32.48	85	2.35	48	2.13	50	0.30	48	31.75	83	1.46	50	1.87	46	0.48	48
S004	35.15	56	2.96	51	2.42	51	0.70	48	32.51	63	2.13	52	2.57	52	1.48	52
S005	22.60	128	1.66	44	2.33	45	0.44	45	22.78	123	1.20	52	1.76	50	0.34	52
S006	-	-	-	-	-	-	-	-	28.32	97	1.53	52	1.77	50	0.23	50
S007	26.38	82	0.99	46	2.59	49	0.54	48	28.20	100	1.86	50	2.09	50	0.54	50
S008	28.24	101	1.56	52	2.60	52	0.50	52	-	-	-	-	-	-	-	-
S009	23.68	44	1.51	44	1.72	44	0.30	44	24.65	38	1.08	38	1.87	37	0.28	37
S010	29.44	78	1.68	47	2.63	46	1.10	46	29.03	81	1.34	52	1.94	52	0.76	52
S011	26.15	43	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S012	28.54	44	1.61	44	2.06	43	2.70	43	-	-	-	-	-	-	-	-
S013	24.67	80	2.06	47	2.20	49	0.88	48	-	-	-	-	-	-	-	-
S014	24.52	19	2.94	17	2.44	17	0.59	17	23.16	20	2.42	19	1.27	19	0.36	19
S015	22.29	24	1.07	21	1.81	21	0.07	21	20.07	35	1.33	35	1.46	33	0.26	34
S016	29.30	80	2.19	51	2.16	51	0.41	51	28.31	81	2.28	51	1.94	50	0.85	49
S017	28.68	45	1.49	45	1.81	45	0.52	45	-	-	-	-	-	-	-	-
S018	24.83	89	2.22	49	1.66	50	0.29	50	-	-	-	-	-	-	-	-
S019	25.75	182	1.62	50	1.96	48	0.28	50	25.27	161	1.11	50	2.13	50	0.43	49
S020	24.34	72	1.44	48	2.10	48	1.09	47	24.40	75	2.65	50	2.15	50	1.10	50
S021	26.39	51	1.79	50	1.68	51	0.18	51	-	-	-	-	-	-	-	-
S022	21.82	80	2.22	49	1.76	49	0.18	49	-	-	-	-	-	-	-	-
S023	29.34	48	2.06	48	1.98	48	0.80	48	30.03	75	1.41	49	1.92	49	0.35	49
S024	26.23	59	2.69	50	1.97	29	1.38	29	-	-	-	-	-	-	-	-
S025	25.99	100	0.81	50	1.28	50	0.21	50	-	-	-	-	-	-	-	-
S026	29.60	90	2.17	3	-	-	-	-	-	-	-	-	-	-	-	-
S027	-	-	-	-	-	-	-	-	26.31	114	1.30	52	1.92	52	0.73	51
S028	25.14	105	1.49	50	2.99	50	1.02	50	26.79	73	1.91	48	2.49	50	1.23	50
S029	26.79	93	2.24	43	1.60	43	0.25	42	-	-	-	-	-	-	-	-
S030	-	-	-	-	-	-	-	-	25.36	85	1.54	51	2.02	50	0.91	50
S031	-	-	-	-	-	-	-	-	24.08	70	0.97	48	1.67	48	0.26	48
S032	18.21	77	0.86	47	1.62	47	0.17	47	15.86	86	0.47	38	1.12	37	0.17	36

School Code	Visual Arts Assessment 2007								Visual Arts Assessment 2008							
	MC 2007	N	Task 1	N	Task 2d	N	Task 2w	N	MC 2008	N	Task 1	N	Task 2d	N	Task 2w	N
S033	15.86	99	1.24	49	1.56	48	0.15	48	-	-	-	-	-	-	-	-
S034	20.84	41	1.51	39	1.63	39	0.00	39	15.86	43	1.92	44	0.92	44	0.08	30
S035	-	-	-	-	-	-	-	-	29.34	98	1.77	48	2.25	51	0.89	50
S036	22.69	30	1.00	31	1.10	31	0.21	29	-	-	-	-	-	-	-	-
S037	-	-	-	-	-	-	-	-	30.51	112	1.87	50	2.59	51	1.09	51
S038	21.63	100	1.07	50	1.64	50	0.06	50	20.92	100	0.55	50	1.53	50	0.22	50
S039	30.13	90	1.98	44	2.43	44	1.47	44	-	-	-	-	-	-	-	-
S040	24.13	64	1.96	48	1.93	28	0.25	28	-	-	-	-	-	-	-	-
S041	26.90	48	2.46	48	2.44	48	2.79	48	28.41	65	3.04	47	2.32	47	2.25	47
S042	30.31	51	1.90	48	2.26	50	0.90	50	-	-	-	-	-	-	-	-
S043	27.12	118	1.65	47	2.12	47	0.27	46	-	-	-	-	-	-	-	-
S044	25.48	82	1.76	48	1.71	28	0.17	26	23.91	90	0.96	52	1.95	53	0.31	51
S045	31.11	74	2.68	50	2.40	51	1.89	51	34.31	71	2.68	51	2.21	50	2.11	50
S046	19.58	63	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S047	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S048	-	-	-	-	-	-	-	-	22.33	67	0.97	48	2.00	47	0.29	47
S049	-	-	-	-	-	-	-	-	24.98	45	1.56	45	1.68	45	0.19	44
S050	23.67	98	1.47	50	1.97	50	1.00	50	20.12	109	1.68	49	1.76	49	1.01	47
S051	27.65	62	0.97	48	2.02	48	0.02	48	-	-	-	-	-	-	-	-
S052	23.45	69	1.56	50	2.08	50	0.40	50	19.62	60	0.74	50	1.41	51	0.11	50
S053	-	-	-	-	-	-	-	-	32.97	102	1.45	42	2.11	42	0.46	42
S054	23.39	111	1.45	49	2.28	50	0.98	50	-	-	-	-	-	-	-	-
S055	19.77	67	1.18	48	1.88	48	0.50	48	-	-	-	-	-	-	-	-
S056	26.04	33	1.55	33	2.00	33	0.41	33	-	-	-	-	-	-	-	-
S057	23.00	26	1.20	25	1.45	10	0.63	12	23.63	28	1.76	28	1.33	27	0.56	27
S058	27.11	58	2.91	37	1.59	41	0.50	40	29.21	68	3.26	51	2.25	51	0.52	51
S059	25.83	81	2.62	50	1.71	48	0.16	49	-	-	-	-	-	-	-	-
S060	21.90	91	2.67	48	2.13	48	1.35	48	23.08	93	2.65	47	2.47	47	1.01	47
S061	25.45	43	1.33	41	2.01	42	0.25	42	23.21	63	1.41	49	1.94	46	0.63	46
S062	21.02	111	1.25	50	2.04	50	0.15	49	-	-	-	-	-	-	-	-
S063	25.41	72	2.15	54	1.66	55	0.46	52	25.43	81	1.75	51	1.64	50	0.60	50
S064	28.03	53	1.64	47	2.04	36	0.93	36	-	-	-	-	-	-	-	-
S065	20.85	68	1.97	51	1.98	52	0.59	51	-	-	-	-	-	-	-	-
S066	33.04	101	2.68	50	2.23	49	1.33	49	-	-	-	-	-	-	-	-

School Code	Visual Arts Assessment 2007								Visual Arts Assessment 2008							
	MC 2007	N	Task 1	N	Task 2d	N	Task 2w	N	MC 2008	N	Task 1	N	Task 2d	N	Task 2w	N
S067	-	-	-	-	-	-	-	-	28.89	101	1.89	52	2.30	51	0.99	51
S068	-	-	-	-	-	-	-	-	29.90	47	2.00	50	1.76	51	0.46	50
S069	18.61	78	1.00	11	2.00	1	0.00	1	-	-	-	-	-	-	-	-
S070	24.17	92	2.06	51	1.94	51	0.97	50	-	-	-	-	-	-	-	-
S071	23.35	47	2.10	47	1.66	47	0.99	47	20.58	60	1.63	47	1.41	45	0.54	46
S072	24.08	74	1.85	50	1.80	50	0.40	50	22.39	67	1.06	50	1.84	50	0.29	50
S073	-	-	-	-	-	-	-	-	32.72	87	1.94	51	2.14	51	0.54	52
S074	25.57	102	2.21	50	1.78	51	0.67	51	-	-	-	-	-	-	-	-
S075	-	-	2.76	46	1.75	46	0.43	46	30.72	99	2.70	50	2.38	50	0.90	49
S076	25.77	91	2.02	49	2.08	49	1.22	49	25.10	90	1.62	51	1.74	50	0.50	50
S077	29.01	102	-	-	-	-	-	-	-	-	-	-	-	-	-	-
S078	23.14	105	1.52	43	2.67	43	0.56	43	-	-	-	-	-	-	-	-
S079	24.84	35	1.29	35	1.99	35	0.27	35	-	-	-	-	-	-	-	-
S080	-	-	-	-	-	-	-	-	27.30	76	0.91	46	1.74	47	0.24	45
S081	24.74	37	1.13	36	1.64	36	0.08	37	-	-	-	-	-	-	-	-
S082	34.11	105	1.98	49	2.35	49	0.65	48	33.03	115	2.39	51	2.42	51	0.58	51
S083	28.43	105	2.19	48	2.21	47	0.45	48	30.81	107	1.82	47	2.14	47	0.85	46
S084	22.69	38	1.72	36	2.08	37	0.49	36	-	-	-	-	-	-	-	-
S085	29.89	93	2.04	46	2.07	46	0.22	46	-	-	-	-	-	-	-	-
S086	22.02	89	0.94	47	2.20	45	0.22	45	19.47	79	1.08	47	1.34	47	0.53	47
S087	-	-	-	-	-	-	-	-	25.46	96	1.11	48	1.67	46	0.35	46
S088	29.27	65	1.97	52	1.85	51	0.60	49	-	-	-	-	-	-	-	-
S089	-	-	-	-	-	-	-	-	19.86	75	0.92	52	1.41	52	0.27	52
S090	19.12	26	1.50	26	1.37	26	0.18	25	25.07	19	2.17	19	1.97	19	2.09	19
S091	25.16	85	1.78	50	2.01	50	0.32	50	-	-	-	-	-	-	-	-
S092	27.61	101	1.32	49	1.95	49	0.15	49	28.73	79	1.33	49	1.95	48	0.34	47
S093	23.08	46	1.75	46	2.36	44	1.14	42	-	-	-	-	-	-	-	-
S094	28.52	53	1.50	17	1.97	15	1.00	15	-	-	-	-	-	-	-	-
S095	22.33	44	1.00	39	1.93	40	0.10	39	-	-	-	-	-	-	-	-
S096	20.10	36	0.67	36	1.53	32	0.35	34	22.04	30	0.78	32	1.28	35	0.10	30
S097	23.12	81	1.49	46	2.28	47	0.31	47	22.41	71	0.73	49	1.46	49	0.14	49
S098	25.56	65	1.45	49	1.96	50	0.24	50	-	-	-	-	-	-	-	-
S099	-	-	-	-	-	-	-	-	29.17	109	1.88	52	2.14	52	0.51	52
S100	-	-	-	-	-	-	-	-	27.74	111	1.39	50	1.72	50	0.54	50

VII. CURRENT YEAR IMPLEMENTATION LOGISTICS

Teachers who administered the 2007-2008 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 15 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, 95.2% of respondents noted that the Test Administrator Training Session adequately prepared them to administer the SCAAP assessment. Several test administrators pointed out that "this year's training session was excellent." Two respondents indicated that they were unable to attend the training session, and one respondent noted feeling unprepared but stated, "[i]t's not anyone's fault. I don't think anyone could have prepared me for all the technical difficulties I had!" Only one respondent offered a critique, writing, "[i]t was almost too much information to learn in such a short amount [of] time."

Regarding technical support and troubleshooting, 100% of the respondents indicated that they were satisfied with the support that they received from the SCAAP personnel, with one respondent stating that the training was "better than ever," and another stating, "I was satisfied with the prompt answers … when the majority of the computers froze." Several respondents wrote positive comments, and, as one participant summarized, "[e]verything was clearly laid out in the manual. I felt that if I had a question that the staff would willingly and quickly answer."

Comments and Suggestions
The test administrators also provided the following comments and suggestions for improving the training session: (1) ensuring that administrators receive the packet for perusal before the training, (2) checking individual school district calendars for conflicts with training sessions, (3) including more hands-on activities during the training session for new teachers, (4) reviewing how to delete students that have taken the test, (5) expanding on the topic "How to read and understand test results," (6) creating separate sessions for new and experienced test administrators or making the training optional for experienced test administrators, (7) proofreading the manual to prevent confusion, (8) ensuring teachers are knowledgeable concerning software, and (9) troubleshooting questions such as, "What happens if a student is knocked offline?, What if a student on the SASI list is handicapped?, and What if the student is new at the school?"

Recommendations and Clarifications
Based on the teacher comments, the SCAAP personnel will make the following improvements to the training sessions and the following support available:
1. Provide test administrators with online access to test administration manuals prior to the training.
2. Create alternative training options for experienced test administrators, such as a web-based training video.

3. Emphasize software usage and adding and deleting students on the web during the training sessions.
4. Include a "Test Administration Frequently Asked Questions" section on the website.
5. Expand the results interpretation section in the test administration manuals.

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 the following issues. The SCAAP personnel should point out to the test administrators that they can request slightly earlier or later testing dates outside the official testing window. As long as the dates are within a reasonable time frame, the SCAAP personnel will work with individual schools to accommodate special requests. Moreover, the SCAAP personnel should emphasize that the only reason for not having additional accommodations for students with specific IEPs in SCAAP testing is financial constraints. Although a few teachers provided valuable suggestions such as "creating a recorded version of the test," the SCAAP assessments are currently not financially equipped to accommodate such requests. Finally, the SCAAP personnel should clarify that they are responsible for the testing aspects of SCAAP, and any requests regarding the DAP grant should be directed to the SCDE.

Administering SCAAP Assessments

A total of 28 test administrators completed the music assessment survey, and 15 test administrators completed the visual arts assessment survey. Of the 28 music assessment respondents, 33.3% also administered the assessment during the 2006-2007 academic year, 25.9% during the 2005-2006 year, and 7.4 during the 2004-2005 year. Of the 15 visual arts assessment respondents, 66.7% also administered the test during the 2006-2007 academic year, 46.7% administered the test during 2005-2006, and 46.7% administered this test during 2004-2005.

Multiple-Choice Assessment
In reference to how long it took to administer the music multiple-choice assessment, 32.1% reported 30 to 45 minutes, 64.3% reported between 46 and 60 minutes, and 3.6% reported over 60 minutes. Of the visual arts respondents, 60% indicated that the multiple-choice assessment took between 30 and 45 minutes, and 40% indicated that it took between 40 and 60 minutes to administer.

In reference to the multiple-choice assessment directions, 92.6% of music test administrators and 85.7% of the visual arts test administrators affirmed that they read the "Test Administrator's Script" exactly as written in the Test Administration Manual when they administered the multiple-choice test. The reasons for not reading the script exactly included having un-trained test administrators proctor the test and having extra computer problems. Although revised copies of the test administration manual for visual arts were sent to the teachers prior to the begin of the official testing dates, a few visual arts teachers still commented that they did not follow the script due to an error in the directions regarding the correct response to a practice test item. One test administrator indicated that he/she did not go through the practice test with the students due to time constraints.

Over 67% of the teachers indicated that they did not have to provide additional instructions to students. Of the 33% who did provide additional instructions, the majority mentioned that adding

instructions to solve computer-related issues such as "be careful about pushing the wrong buttons," and "go back and read and check answers." Test administrators also explained to students how to skip and return to specific questions, and told students they could listen to the musical examples more than once. In addition, instructions related to classroom management were often included such as "raise your hand if you have a question." One teacher indicated that she added "take your time, click once, put on your headsets, do your best, raise your hand if you have a question, and 'You are ARTS SMART!'" in her instruction. Other teachers indicated reading specific question to students with special needs.

On average, about 80% of the 28 respondents reported that at least one computer froze up during testing. A few respondents indicated that the graphics were not displayed and sounds did not play. One school mentioned that the problem with sounds may not only be related to bandwidth limitation but also to the choice of media player for playing embedded sounds in Internet Explorer. They suggested not using QuickTime as a preferred media player as they encountered extensive problems with this software. One test administrator noted that they experienced difficulties due to old headphones.

Several test administrators indicated that they were unaware that they could request for special codes that allow students to return to the middle of the test due to emergency, interruption or technology related problems. One test administrator commented that

> *There really needs to be a way to store a student's test so they [sic] can return to it in case of an emergency interruption ... The only thing we can do now is delete results and have them start all over again, or not finish the test. We should be able to save the questions answered and let the student log back in and start where they [sic] left off. You could set it up where they could not go back and change any questions they answered ... if you are worried about test security.*

SCAAP personnel do not provide the special codes for getting back to the middle of tests with the actual test codes to avoid schools dividing the test into several parts. During normal school hours, the 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 has been repeatedly discussed during the test administration training sessions, the SCAAP personnel agreed that this information should be included in the test administration manual as a separate section and on the FAQ section on the SCAAP website.

Music Performance Task Assessment
About 60% of the test administrator reported that they used less than five minutes and about 30% reported that they used between 5 and 10 minutes to administer each music performance task to each student. About 86% of the test administrators reported that the student understood the performance tasks directions without problem. One test administrator indicated that his/her students experienced more difficulty with the singing task while another teacher reported that his/her students experienced more difficulty with the improvisation task. Several teachers indicated that their students had difficulty knowing when to begin their rhythm improvisation. Although the example of when a student should begin performing was repeated four times in the directions, a few teachers felt that "it was not emphasized enough."

Specific suggestions were also provided regarding the content of the performance task scripts and the performance task format such as (a) providing written instructions in addition to the CD, (b) slowing down the tempo for the improvisation task, and (c) changing the order of the instructions. Those suggestions will be presented to the SCAAP Music Advisory Committee at this year's item review session.

Finally, scheduling and the amount of time required to administer the test was the most reported problem. Teachers included three main recommendations for decreasing the problems associated with scheduling and administration time. Some teachers requested that school principals be informed regarding the amount of time required for administering the music performance tasks. The SCAAP personnel recommend adding specific information regarding the time required for test administration to the SCAAP Participation Agreement form included with the DAP grant application. Second, teachers suggested that schools be permitted to hire substitute teachers to assist in administering the performance tasks. Although schools are currently permitted to allocate a portion of the DAP grant funds to pay for substitute teachers to be used during SCAAP test administration, many participating teachers are not aware of that possibility exists. The SCAAP personnel recommend increasing the visibility of that feature of the DAP grant so that more teachers will take advantage of the provided accommodation. Finally, some teachers recommend requiring a smaller sample to take the test. In 2009, the amount of students who are required to take the performance tasks has been decreased from 50 students to 40 students.

Visual Arts Performance Task Assessment
The majority of test administrators reported that they took less than 30 minutes to administer each Visual Arts Performance Task. A few of them indicated that they spent less than 10 minutes on each task. About 87% of the respondent indicated that student understood the performance task directions and 67% did not have to provide additional instructions. In explaining why students did not understand the directions, one respondent stated, "I had several questions about whether they had to use all of the words in the word bank or just some. I reread the instruction[s] and told them that they would have to interpret what they thought." Another respondent explained,

> They really have a hard time with explaining where things are in the picture. They assume that if they see something (colors, lines, etc.) in a certain area that the graders see the same thing. I am not saying that they shouldn't work on this. We have lots of oral discussions in class and students have so many good things to say. When they start to write, they stumble over spelling and grammar. Maybe they should do like they do on the music portion and record their answers. We will continue to work on this area with classroom teachers.

Additional instructions provided by test administrators included re-reading instructions, explaining terms used in the instructions such as "compare and contrast," answering questions, and giving additional details such as "you can turn your paper sideways" and "make sure your drawing is dark enough to show up on a computer scan."

The survey also asked test administrators their suggestions for improving the implementation of the performance tasks. The suggestions included the following: (1) provide detailed instructions on what can and cannot be discussed with the students about the questions or tasks, (2) provide a formal script for the performance assessment, (3) provide "an example of the types of answers you desire," and (4) suggest that principals allow a place and time to conduct this portion of the

test. As one respondent explained, "I gave it over the course of about three weeks. Students need time to plan what they are going to write. They also need about 45 minutes to draw a good picture and more time to answer the two questions. They are fourth graders."

Using Assessment Results

Test administrators also explained how they planned to use the results from the SCAAP assessment. Overall, responses indicated that the assessments were useful for improving individual instruction and for school-wide curriculum planning. 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 and "concepts that need to be stressed more rigorously in future lesson plans." Several respondents further specified that the assessments are also used "to determine strengths and weaknesses of our arts program" and "to determine which areas of [the] art program need to be redesigned in terms of year-long planning." Three respondents mentioned that the assessments are used for comparisons with other schools, and one respondent mentioned that the tests are used for grants.

Only two respondents provided negative responses in relation to the SCAAP assessments. One respondent wrote, "[w]ell, the multi-choice was a mess technically, so I'm not even sure the results will really reflect what they know." Another respondent replied with the following critique:

> I don't like the way you give results. "This is the percentage of students who scored well on these standards." Some of the questions that are on the test to cover some standard are out in left field. You are making a statement about the strengths and weaknesses of my students based on two or three obscure questions. Every year there are more strange questions. I wish you would look at the level of questioning in fourth-grade art books. I don't really get the purpose of this test.

Those negative comments indicated to the SCAAP personnel that additional explanation regarding the development of the SCAAP assessments is needed to clarify arts educators' misunderstanding of the implementation of the assessments. 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. One respondent mentioned experiencing scheduling difficulties due to limited availability of the computer lab. One music teacher expressed concern that the multiple-choice was too, "heavy on listening and instruments." Two respondents noted that the size of the flash drives is not sufficient to hold all the sound files. Furthermore, two respondents noted that not having to scan student responses was very helpful. One respondent thanked SCAAP staff "for not requiring us to upload the performance tasks," and another thanked SCAAP "for registering the student passwords and usernames. That was so helpful!"

One respondent wrote, "I do like having the feedback. We see our students only once a week and therefore only really assess them by observation." Finally, two respondents suggested expanding the program to all schools. One stated, "I think it is a great opportunity for student[s] and teachers to see where we are with meeting and teaching the standards. Thank you."

Test administrators suggested the following future changes to the assessment: (1) change the scheduling to after/before PACT or at the beginning of students' fifth-grade year, (2) ensure that questions are standards-based and not curriculum-based, (3) notify teachers of particular artists to be covered on the assessment, (4) accommodate the needs of students with IEPs and new students, (5) shorten the assessments, (6) provide results earlier, and (7) align tests to South Carolina or district curriculum.

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.

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

SDE (1999). *South Carolina Visual and Performing Arts Standards*. SC: South Carolina Department of Education.

SDE (2002). *South Carolina Visual and Performing Arts Curriculum Standards 2002 Field Review Draft*. 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: 2008 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	0.84	0.33	83*	4	6	5	A	A
2	0.80	0.38	11	80*	4	3	A	B
3	0.73	0.35	11	72*	7	7	A	A
4	0.46	0.35	15	10	28	45*	A	A
5	0.72	0.39	72*	12	4	10	A	B
6	0.38	0.30	15	38*	28	17	A	A
7	0.69	0.28	69*	10	10	10	A	A
8	0.79	0.35	2	13	5	78*	A	A
9	0.51	0.28	51*	7	18	22	A	A
10	0.39	0.40	11	38*	24	25	A	A
11	0.55	0.44	26	5	54*	12	A	A
12	0.49	0.45	49*	11	14	24	A	A
13	0.77	0.39	4	3	76*	15	A	A
14	0.47	0.34	11	47*	23	17	A	A
15	0.27	0.28	34	7	31	26*	A	A
16	0.36	0.32	36*	26	21	16	A	A
17	0.37	0.31	34	36*	13	14	A	A
18	0.67	0.34	10	66*	12	10	A	A
19	0.53	0.40	12	13	53*	21	A	A
20	0.44	0.30	12	22	20	44*	A	A
21	0.44	0.35	3	14	44*	11	B	A
22	0.45	0.34	18	45*	17	18	A	A
23	0.65	0.49	64*	14	11	8	A	A
24	0.48	0.36	48*	19	18	13	A	A
25	0.13	0.07	60	14	12	12*	A	A
26	0.31	0.44	29	32	7	30*	A	A
27	0.73	0.37	3	73*	2	20	A	A
28	0.51	0.47	12	21	15	51*	A	A
29	0.92	0.32	2	2	2	93*	A	A
30	0.62	0.26	3	28	5	62*	A	C
31	0.34	0.32	34*	7	29	27	A	A
32	0.35	0.27	17	14	34*	32	A	A
33	0.34	0.29	21	34*	20	23	A	A
34	0.50	0.29	9	23	16	50*	A	A
35	0.56	0.21	17	56*	14	11	A	A
36	0.48	0.22	49*	15	8	26	A	A
37	0.60	0.39	21	6	11	60*	A	A
38	0.67	0.42	68*	11	11	8	A	A
39	0.40	0.28	40*	29	15	14	A	A
40	0.43	0.34	24	44*	15	16	A	A
41	0.56	0.33	56*	15	11	16	A	A
42	0.46	0.35	46*	24	8	19	A	A
43	0.60	0.40	10	7	61*	19	A	A
44	0.66	0.42	14	6	12	66*	A	A
45	0.86	0.41	3	87*	4	4	A	B

APPENDIX B: 2008 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	0.83	0.34	83*	4	6	5	B	A
2	0.65	0.41	11	11	65*	11	A	A
3	0.77	0.35	11	76*	3	8	A	A
4	0.58	0.31	7	12	22	57*	B	A
5	0.47	0.37	12	13	27	47*	A	A
6	0.66	0.25	15	66*	12	5	A	A
7	0.39	0.22	12	39*	31	16	A	A
8	0.34	0.41	10	34*	25	29	A	A
9	0.42	0.47	34	5	42*	17	A	A
10	0.76	0.36	3	15	5	75*	A	A
11	0.39	0.31	13	38*	42	5	A	A
12	0.53	0.47	27	5	53*	14	A	A
13	0.38	0.46	38*	19	10	32	A	B
14	0.50	0.32	11	50*	22	15	A	A
15	0.36	0.26	36*	28	25	10	A	A
16	0.49	0.31	48*	10	19	21	A	A
17	0.61	0.41	16	19	2	60*	A	A
18	0.70	0.34	7	69*	10	12	A	A
19	0.43	0.29	16	42*	18	21	A	A
20	0.44	0.31	16	18	21	43*	A	A
21	0.54	0.42	10	15	54*	19	A	A
22	0.46	0.36	45*	18	20	15	A	A
23	0.56	0.43	22	56*	13	7	A	A
24	0.18	0.15	50	16	15	17*	A	A
25	0.39	0.29	18	38*	18	23	A	A
26	0.61	0.41	61*	17	7	13	A	A
27	0.63	0.16	13	63*	10	13	A	B
28	0.37	0.29	21	27	37*	14	A	A
29	0.81	0.39	6	4	81*	8	A	A
30	0.65	0.37	10	7	65*	16	A	A
31	0.65	0.43	14	8	11	65*	A	A
32	0.94	0.31	1	1	1	94*	B	A
33	0.32	0.31	31*	8	31	27	A	A
34	0.57	0.22	5	19	17	57*	A	A
35	0.41	0.22	41*	28	14	15	A	A
36	0.42	0.33	21	42*	15	20	A	A
37	0.40	0.44	21	27	11	39*	A	A
38	0.44	0.45	20	44*	16	18	A	A
39	0.45	0.41	16	45*	15	23	A	A
40	0.49	0.24	49*	17	8	24	A	A
41	0.72	0.51	10	8	7	72*	A	A
42	0.59	0.45	20	6	12	59*	A	A
43	0.55	0.39	22	55*	13	8	A	A
44	0.87	0.39	3	88*	3	4	A	A
45	0.74	0.39	11	75*	4	7	A	A

APPENDIX C: 2008 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	0.81	0.31	2	12	4	80*	A	A
2	0.68	0.45	6	13	12	67*	A	B
3	0.74	0.43	11	9	73*	5	A	A
4	0.67	0.38	12	67*	6	13	A	A
5	0.74	0.41	6	15	2	74*	A	A
6	0.74	0.23	8	5	10	75*	A	A
7	0.82	0.28	6	7	81*	4	A	A
8	0.67	0.43	11	13	8	66*	A	A
9	0.70	0.43	70*	11	12	4	A	A
10	0.65	0.34	10	64*	14	9	A	A
11	0.79	0.29	79*	11	4	5	A	A
12	0.17	0.00	21	25	34	17*	A	A
13	0.55	0.19	20	18	5	54*	A	A
14	0.69	0.42	69*	10	14	6	A	A
15	0.67	0.43	21	5	6	66*	A	A
16	0.73	0.30	7	7	73*	11	A	A
17	0.70	0.50	4	19	69*	5	B	A
18	0.15	0.15	15*	34	10	39	A	A
19	0.42	0.39	19	20	18	42*	A	A
20	0.50	0.42	27	50*	17	4	A	A
21	0.77	0.47	77*	11	5	5	A	B
22	0.41	0.29	10	40*	24	24	A	A
23	0.68	0.44	7	8	68*	16	A	A
24	0.44	0.30	43*	10	29	15	A	A
25	0.31	0.27	23	22	30*	23	A	A
26	0.58	0.42	16	11	13	58*	A	A
27	0.63	0.50	4	63*	7	24	A	A
28	0.58	0.35	8	9	57*	24	A	B
29	0.74	0.33	7	73*	6	12	A	A
30	0.42	0.31	21	16	20	42*	A	A
31	0.65	0.51	9	9	15	65*	A	A
32	0.43	0.28	42*	22	16	18	A	A
33	0.21	0.06	55	10	21*	12	A	A
34	0.30	0.24	39	19	10	29*	A	A
35	0.61	0.34	18	61*	6	13	A	A
36	0.60	0.44	60*	21	10	7	A	A
37	0.25	0.32	37	24*	23	14	A	A
38	0.52	0.51	8	51*	11	27	A	A
39	0.24	0.18	30	31	12	24*	A	A
40	0.33	0.03	6	44	33*	15	A	A
41	0.42	0.41	18	41*	15	24	A	A
42	0.58	0.47	14	10	16	58*	A	A
43	0.25	0.03	22	25*	32	18	A	B
44	0.32	0.39	32*	15	36	14	A	A
45	0.55	0.48	20	13	10	55*	A	A

APPENDIX D: 2008 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	0.84	0.25	7	4	4	83*	A	A
2	0.69	0.46	6	12	12	69*	A	A
3	0.73	0.41	9	9	73*	6	A	A
4	0.70	0.35	10	70*	5	14	A	A
5	0.74	0.39	7	15	2	74*	A	A
6	0.75	0.22	8	5	9	76*	A	A
7	0.81	0.43	4	3	80*	10	A	A
8	0.66	0.25	11	13	8	66*	A	A
9	0.69	0.41	68*	12	13	5	A	A
10	0.62	0.38	13	61*	13	11	A	A
11	0.61	0.44	8	16	14	60*	B	A
12	0.19	-0.02	23	24	33	19*	A	A
13	0.48	0.33	20	2	48*	28	A	A
14	0.65	0.42	9	64*	12	13	A	A
15	0.67	0.44	20	5	6	67*	A	A
16	0.79	0.41	4	5	79*	11	A	A
17	0.66	0.48	5	22	65*	5	A	A
18	0.16	0.12	16*	39	15	27	A	A
19	0.41	0.42	20	22	16	41*	A	A
20	0.46	0.39	30	46*	11	12	A	A
21	0.48	0.41	27	48*	14	10	A	A
22	0.49	0.42	49*	10	16	23	A	A
23	0.73	0.43	12	73*	8	5	A	A
24	0.53	0.57	11	21	13	52*	A	A
25	0.31	0.18	32	31*	19	16	A	A
26	0.54	0.43	21	11	12	54*	A	A
27	0.63	0.49	4	63*	7	24	A	A
28	0.53	0.39	9	10	53*	25	A	A
29	0.78	0.40	5	78*	5	10	A	A
30	0.48	0.33	8	11	31	48*	B	A
31	0.66	0.51	9	8	15	66*	A	A
32	0.43	0.26	24	43*	15	17	A	A
33	0.24	0.04	51	12	24*	11	A	A
34	0.31	0.23	36	20	12	30*	A	A
35	0.62	0.45	8	17	62*	11	A	A
36	0.55	0.44	55*	23	12	8	A	A
37	0.21	0.25	37	20*	20	20	A	A
38	0.53	0.44	17	14	15	53*	A	A
39	0.39	0.32	17	26	17	38*	A	A
40	0.34	0.02	6	42	34*	17	A	A
41	0.42	0.37	17	42*	15	24	A	A
42	0.55	0.47	14	12	18	55*	A	A
43	0.63	0.18	9	8	63*	19	A	A
44	0.33	0.22	15	33	17	33*	A	A
45	0.59	0.47	18	11	11	59*	A	A

APPENDIX E: 2008 SCAAP Item P-Values for Music Test Forms by Gender

	Music Form 1			Music Form 2		
Item	Female	Male	ETS Δ	Female	Male	ETS Δ
1	0.86	0.81	0.65	0.87	0.80	1.02
2	0.83	0.78	0.57	0.69	0.62	0.54
3	0.77	0.69	0.62	0.81	0.73	0.83
4	0.46	0.46	-0.33	0.64	0.52	1.10
5	0.76	0.69	0.58	0.49	0.45	0.17
6	0.42	0.34	0.48	0.65	0.67	-0.45
7	0.72	0.67	0.40	0.41	0.37	0.30
8	0.80	0.77	0.12	0.35	0.34	-0.14
9	0.54	0.48	0.37	0.42	0.43	-0.38
10	0.40	0.37	-0.05	0.78	0.74	0.31
11	0.53	0.56	-0.89	0.36	0.42	-0.88
12	0.48	0.50	-0.70	0.52	0.55	-0.77
13	0.77	0.76	-0.36	0.37	0.40	-0.82
14	0.46	0.48	-0.53	0.50	0.51	-0.31
15	0.30	0.24	0.45	0.36	0.36	-0.21
16	0.36	0.36	-0.25	0.50	0.47	0.08
17	0.38	0.35	0.08	0.59	0.62	-0.57
18	0.70	0.63	0.51	0.74	0.66	0.77
19	0.52	0.54	-0.64	0.44	0.42	-0.04
20	0.47	0.42	0.23	0.48	0.40	0.56
21	0.40	0.48	-1.24	0.54	0.55	-0.43
22	0.46	0.44	-0.16	0.50	0.42	0.56
23	0.68	0.61	0.45	0.57	0.55	-0.06
24	0.51	0.45	0.32	0.18	0.18	-0.16
25	0.13	0.13	-0.16	0.40	0.38	0.07
26	0.30	0.31	-0.62	0.65	0.57	0.60
27	0.74	0.72	-0.05	0.65	0.61	0.26
28	0.54	0.48	0.24	0.38	0.37	-0.13
29	0.94	0.91	0.21	0.82	0.80	0.01
30	0.68	0.57	0.87	0.66	0.64	0.07
31	0.36	0.32	0.14	0.64	0.66	-0.63
32	0.36	0.33	0.02	0.96	0.92	1.24
33	0.37	0.31	0.41	0.31	0.32	-0.39
34	0.51	0.49	-0.10	0.60	0.54	0.47
35	0.58	0.54	0.30	0.39	0.42	-0.52
36	0.51	0.46	0.28	0.41	0.44	-0.58
37	0.63	0.57	0.25	0.40	0.39	-0.31
38	0.69	0.66	-0.06	0.46	0.42	0.08
39	0.39	0.41	-0.44	0.46	0.44	-0.05
40	0.43	0.44	-0.43	0.49	0.50	-0.29
41	0.57	0.55	-0.02	0.73	0.70	0.14
42	0.49	0.43	0.27	0.60	0.58	-0.05
43	0.61	0.60	-0.26	0.56	0.53	-0.01
44	0.64	0.67	-0.77	0.87	0.86	-0.30
45	0.86	0.85	-0.49	0.78	0.70	0.76

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

APPENDIX F: 2008 SCAAP Item P-Values for Music Test Forms by Ethnicity

Item	Music Form 1 African-American	White	ETS Δ	Music Form 2 African-American	White	ETS Δ
1	0.80	0.87	0.60	0.79	0.87	0.51
2	0.71	0.88	-1.07	0.58	0.73	0.12
3	0.67	0.77	0.44	0.72	0.82	0.29
4	0.39	0.51	0.14	0.53	0.63	0.36
5	0.70	0.76	1.24	0.39	0.55	-0.20
6	0.34	0.40	0.58	0.61	0.72	-0.26
7	0.66	0.72	0.68	0.36	0.42	0.51
8	0.73	0.84	0.12	0.26	0.41	-0.02
9	0.47	0.55	0.34	0.34	0.50	0.45
10	0.30	0.46	-0.18	0.67	0.83	-0.58
11	0.43	0.63	-0.38	0.32	0.46	-0.17
12	0.41	0.56	0.31	0.42	0.64	-0.29
13	0.69	0.82	0.00	0.24	0.51	-1.17
14	0.41	0.52	0.18	0.45	0.55	0.29
15	0.22	0.31	-0.02	0.32	0.39	0.44
16	0.28	0.42	-0.45	0.43	0.55	0.15
17	0.31	0.41	0.15	0.51	0.69	-0.21
18	0.59	0.72	-0.04	0.61	0.77	-0.47
19	0.40	0.62	-0.85	0.35	0.51	-0.43
20	0.37	0.49	-0.02	0.41	0.47	0.91
21	0.34	0.52	-0.56	0.44	0.63	-0.13
22	0.37	0.51	-0.09	0.41	0.50	0.74
23	0.51	0.76	-0.93	0.45	0.66	-0.49
24	0.43	0.52	0.64	0.17	0.18	0.75
25	0.13	0.13	0.45	0.32	0.44	0.05
26	0.19	0.39	-0.73	0.51	0.71	-0.31
27	0.65	0.79	-0.12	0.64	0.62	1.07
28	0.44	0.57	0.79	0.30	0.44	-0.37
29	0.89	0.95	0.27	0.75	0.86	0.27
30	0.63	0.62	1.61	0.56	0.73	-0.29
31	0.28	0.38	0.15	0.53	0.75	-0.69
32	0.28	0.39	-0.28	0.91	0.97	-0.25
33	0.28	0.38	-0.10	0.27	0.36	0.43
34	0.45	0.53	0.40	0.52	0.61	0.00
35	0.54	0.57	0.59	0.36	0.45	-0.06
36	0.45	0.51	0.38	0.33	0.51	-0.55
37	0.50	0.67	-0.23	0.27	0.51	-0.74
38	0.57	0.76	-0.50	0.32	0.55	-0.52
39	0.32	0.46	-0.41	0.37	0.52	0.28
40	0.33	0.51	-0.58	0.46	0.51	0.60
41	0.45	0.65	-0.81	0.61	0.81	-0.15
42	0.39	0.52	0.13	0.49	0.68	-0.13
43	0.51	0.67	-0.10	0.47	0.61	0.34
44	0.54	0.74	-0.69	0.80	0.92	0.07
45	0.76	0.93	-1.34	0.64	0.81	-0.41

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

APPENDIX G: 2008 SCAAP Item P-Values for Visual Arts Test Forms by Gender

	Visual Arts Form 1		.	Visual Arts Form 2		.
Item	Female	Male	ETS Δ	Female	Male	ETS Δ
1	0.80	0.82	-0.96	0.85	0.83	-0.05
2	0.70	0.66	-0.08	0.70	0.68	-0.28
3	0.77	0.71	0.37	0.77	0.70	0.47
4	0.66	0.68	-0.80	0.70	0.71	-0.57
5	0.78	0.71	0.49	0.77	0.71	0.44
6	0.77	0.70	0.68	0.78	0.71	0.71
7	0.83	0.80	0.04	0.80	0.81	-0.74
8	0.72	0.62	0.67	0.68	0.64	0.24
9	0.74	0.67	0.33	0.72	0.65	0.36
10	0.65	0.64	-0.27	0.61	0.63	-0.63
11	0.80	0.78	-0.25	0.67	0.54	1.03
12	0.17	0.18	-0.27	0.16	0.22	-0.97
13	0.57	0.53	0.19	0.50	0.46	0.08
14	0.69	0.69	-0.63	0.69	0.60	0.63
15	0.68	0.65	-0.17	0.69	0.65	-0.11
16	0.74	0.72	-0.11	0.82	0.75	0.70
17	0.76	0.64	1.06	0.69	0.63	0.25
18	0.15	0.15	-0.34	0.16	0.16	-0.15
19	0.41	0.44	-0.78	0.41	0.41	-0.30
20	0.51	0.49	-0.23	0.47	0.45	-0.07
21	0.80	0.75	0.08	0.50	0.46	0.02
22	0.41	0.40	-0.21	0.53	0.45	0.52
23	0.72	0.64	0.53	0.73	0.74	-0.68
24	0.47	0.40	0.41	0.53	0.53	-0.75
25	0.36	0.26	0.98	0.31	0.32	-0.23
26	0.64	0.53	0.84	0.57	0.52	0.14
27	0.63	0.63	-0.58	0.64	0.63	-0.55
28	0.61	0.54	0.35	0.55	0.51	0.11
29	0.77	0.70	0.49	0.81	0.75	0.48
30	0.40	0.44	-0.77	0.44	0.52	-1.11
31	0.71	0.59	0.92	0.71	0.61	0.71
32	0.44	0.41	-0.09	0.43	0.43	-0.20
33	0.23	0.19	0.51	0.27	0.21	0.68
34	0.28	0.31	-0.65	0.30	0.31	-0.29
35	0.60	0.62	-0.65	0.64	0.60	0.04
36	0.62	0.58	-0.19	0.56	0.54	-0.29
37	0.25	0.24	-0.36	0.19	0.23	-0.90
38	0.53	0.50	-0.42	0.57	0.48	0.53
39	0.25	0.23	0.04	0.39	0.38	-0.17
40	0.33	0.34	-0.15	0.34	0.34	-0.16
41	0.42	0.41	-0.34	0.43	0.41	-0.09
42	0.62	0.53	0.47	0.58	0.51	0.40
43	0.25	0.26	-0.06	0.67	0.58	0.68
44	0.34	0.31	-0.04	0.33	0.33	-0.19
45	0.58	0.52	0.12	0.60	0.57	-0.03

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

APPENDIX H: 2008 SCAAP Item P-Values for Visual Arts Forms by Ethnicity

	Visual Arts Form 1			Visual Arts Form 2		
Item	African-American	White	ETS Δ	African-American	White	ETS Δ
1	0.76	0.86	-0.16	0.80	0.87	0.09
2	0.62	0.73	1.09	0.59	0.77	0.05
3	0.63	0.82	-0.50	0.63	0.82	-0.62
4	0.60	0.73	0.28	0.63	0.77	-0.10
5	0.65	0.82	-0.31	0.66	0.80	0.09
6	0.72	0.76	0.63	0.73	0.76	0.76
7	0.76	0.86	-0.26	0.71	0.89	-0.88
8	0.58	0.74	0.16	0.61	0.71	-0.17
9	0.62	0.77	0.11	0.62	0.75	0.30
10	0.57	0.70	0.00	0.52	0.70	-0.26
11	0.73	0.84	-0.25	0.54	0.67	0.80
12	0.19	0.16	0.74	0.22	0.17	0.83
13	0.53	0.57	0.46	0.39	0.57	-0.41
14	0.61	0.76	0.02	0.56	0.72	0.08
15	0.59	0.73	0.42	0.58	0.75	0.06
16	0.68	0.78	0.04	0.72	0.85	0.02
17	0.57	0.81	-0.85	0.52	0.77	-0.93
18	0.12	0.18	-0.17	0.16	0.17	0.71
19	0.34	0.49	0.20	0.32	0.48	0.40
20	0.38	0.61	-0.71	0.40	0.51	0.81
21	0.64	0.88	-1.55	0.34	0.59	-0.94
22	0.35	0.46	0.26	0.39	0.57	0.07
23	0.59	0.76	0.08	0.64	0.82	-0.45
24	0.38	0.49	0.15	0.36	0.67	-0.73
25	0.25	0.35	0.14	0.27	0.35	-0.27
26	0.51	0.66	0.27	0.46	0.61	0.49
27	0.51	0.72	-0.11	0.51	0.74	-0.36
28	0.44	0.68	-1.11	0.41	0.64	-0.87
29	0.68	0.78	0.28	0.70	0.85	-0.50
30	0.33	0.50	-0.55	0.41	0.54	0.09
31	0.55	0.73	0.20	0.57	0.74	0.61
32	0.38	0.45	0.60	0.35	0.51	-0.66
33	0.18	0.23	-0.47	0.22	0.25	-0.27
34	0.24	0.34	-0.06	0.24	0.36	-0.50
35	0.53	0.67	0.11	0.51	0.71	-0.06
36	0.49	0.69	-0.22	0.45	0.65	-0.17
37	0.17	0.30	-0.22	0.16	0.24	0.00
38	0.39	0.62	-0.26	0.43	0.60	0.46
39	0.19	0.28	-0.32	0.34	0.42	0.73
40	0.37	0.32	0.83	0.38	0.31	0.84
41	0.31	0.50	-0.22	0.33	0.49	0.10
42	0.48	0.66	0.13	0.44	0.64	0.12
43	0.29	0.23	1.00	0.61	0.64	0.57
44	0.24	0.39	0.24	0.29	0.36	0.36
45	0.41	0.67	-0.77	0.46	0.69	-0.32

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