



**South Carolina Arts Assessment Program 2013**  
***Technical Report Prepared for the South Carolina Department of Education***

Submitted October 23, 2013

Xiaofang Zhang, Program Coordinator  
Dr. Ashlee Lewis, Program Director  
Whitney Smiley, Research Assistant  
Yin Burgess, Research Assistant  
Dr. Tammiee Dickenson, Technical Advisor

## TABLE OF CONTENTS

List of Tables .....	iv
List of Figures .....	v
I. INTRODUCTION .....	1
SCAAP Purpose and Goals .....	1
SCAAP Collaborators .....	1
Dissemination & Research .....	6
II. TEST FRAMEWORK & CONTENT .....	7
SCAAP Format and Test Items .....	7
Stimulus Material .....	7
Item Review .....	7
SCAAP Website .....	9
Music Assessment Content & Test Specifications .....	10
SC Music Standards and Indicators .....	10
SCAAP Music Test Specifications .....	14
Music Multiple-Choice Section: Format & Scoring .....	14
Music Performance Task Section: Format & Scoring .....	15
SC Visual Arts Standards and Indicators .....	19
SCAAP Visual Arts Test Specifications .....	21
Visual Arts Performance Task Section: Format & Scoring .....	22
III. TEST ADMINISTRATION .....	27
Participants .....	27
Training Test Administrators .....	27
Administration Procedures .....	28
Multiple-Choice Administration: Music & Visual Arts .....	28
Performance Task Administration: Music .....	28
Performance Task Administration: Visual Arts .....	29

IV. SCORING .....	30
Multiple-choice section .....	30
Performance Task Section .....	30
Benchmarking Music and Visual Arts Performance Tasks .....	30
Web-based Rating System and Procedure .....	30
V. ANALYSIS & RESULTS .....	33
Multiple-Choice Section .....	33
Analysis .....	33
Reliability .....	33
Descriptive Statistics .....	36
IRT Model Fit .....	37
Equating Test Forms .....	38
Technical Characteristics of Items .....	38
Performance Task Section .....	47
Inter-Rater Reliability of Performance Tasks .....	47
Descriptive Statistics .....	49
Task Difficulty Index and Discrimination Index .....	51
Internal Validity .....	52
VI. RESULTS BY SCHOOL .....	55
VII. TEST ADMINISTRATOR FEEDBACK SURVEY .....	60
Demographics of Survey Respondents .....	60
Test Administrator Training Procedures .....	60
Administering the SCAAP Multiple-Choice Assessments .....	61
Administering the SCAAP Performance Task Assessments .....	63
The Music Performance Tasks .....	63
The Visual Arts Performance Tasks .....	63
VIII. CONCLUSIONS .....	66

Technical Characteristics ..... 66

Student Performance ..... 66

Highlights of 2012-2013 SCAAP Development..... 67

References ..... 68

Appendix A: 2012 SCAAP BILOG Syntax..... 71

Appendix B: Supplemental Tables ..... 72



## LIST OF TABLES

Table 1.1: <i>List of SCAAP Advisory Committee Members for the 2013 Assessment Year</i> .....	3
Table 2.1: <i>Table of Specifications for SCAAP Music Assessment</i> .....	14
Table 2.2 : <i>Percent of Items on Music Multiple-Choice Forms by Standard</i> .....	15
Table 2.3: <i>SCAAP 2013 Music Task 1 Rubric—Tonal Criteria</i> .....	16
Table 2.4 : <i>SCAAP 2013 Music Task 1 Rubric—Rhythm Criteria</i> .....	17
Table 2.5 : <i>SCAAP 2013 Music Task 1 Rubric—Vocal Quality Criteria</i> .....	17
Table 2.6: <i>SCAAP 2013 Music Task 2 Rubric—Rhythm Criteria</i> .....	17
Table 2.7: <i>SCAAP 2013 Music Task 2 Rubric—Improvisation Criteria</i> .....	18
Table 2.8: <i>Maximum Points for Music Assessment Performance Tasks</i> .....	18
Table 2.9: <i>Table of Specifications for SCAAP Visual Arts Assessment</i> .....	21
Table 2.10: <i>Percent of Items on Visual Arts Multiple-Choice Forms by Standard</i> .....	22
Table 2.11: <i>SCAAP 2013 Visual Arts Task 1 Rubric—Compare/Contrast</i> .....	24
Table 2.12: <i>SCAAP 2013 Visual Arts Task 2 Rubric—Drawing</i> .....	25
Table 2.13: <i>Maximum Points for Visual Arts Performance Tasks</i> .....	26
Table 5.1: <i>SCAAP 2013 Reliability Indices for Music and Visual Arts Multiple-Choice Items</i> .....	34
Table 5.2: <i>SCAAP 2013 Reliability Indices for Music Multiple-Choice of “Understanding and Listening” sections</i> .....	35
Table 5.3: <i>SCAAP 2013 Reliability Indices for Music Multiple-Choice Assessment by SC Standard</i> .....	35
Table 5.4: <i>SCAAP 2013 Reliability Indices for Visual Arts Multiple-Choice Assessment by SC Standard</i> .....	36
Table 5.5: <i>SCAAP 2013 Descriptive Statistics for Music and Visual Arts Test Forms</i> .....	37
Table 5.6: <i>SCAAP 2013 Ranges of P-values and Discrimination Indices of Test Forms</i> .....	40
Table 5.7: <i>SCAAP 2013 Number of Items with Low and Negative Discrimination Indices</i> .....	41
Table 5.8: <i>SCAAP 2013 Descriptive Statistics of Test Scores by Gender</i> .....	42
Table 5.9: <i>SCAAP 2013 Descriptive Statistics of Test Scores by Ethnicity</i> .....	42
Table 5.10: <i>SCAAP 2013 Ranges of P-values of Test Forms by Gender and Ethnic Groups</i> .....	43
Table 5.11: <i>SCAAP 2013 Summary of DIF Classification for Multiple-Choice Test Forms</i> .....	46
Table 5.12: <i>SCAAP 2013 Inter-Rater Reliability of Music Performance Tasks Using Analytic Rubrics</i> .....	48
Table 5.13: <i>SCAAP 2013 Inter-Rater Reliability of Visual Arts Performance Tasks Using Holistic Rubrics</i> .....	48

Table 5.14: SCAAP 2013 Descriptive Statistics for Music Performance Tasks Using Analytic Rubrics .....	50
Table 5.15: SCAAP 2013 Descriptive Statistics for Visual Arts Performance Tasks Using Holistic Rubrics .....	51
Table 5.16: SCAAP 2013 Task Difficulty Indices and Discrimination Indices for Music and Visual Arts Performance Tasks .....	51
Table 5.17: SCAAP 2013 Correlations for Music Multiple-Choice Test Forms and Performance Tasks.....	53
Table 5.18: SCAAP 2013 Correlations for Visual Arts Multiple-Choice Test Forms and Performance Tasks.....	54
Table 6.1: SCAAP 2013 Music Assessment Means by School.....	56
Table 6.2: SCAAP 2013 Visual Arts Assessment Means by School.....	58
Table 7.1: Test Administrators' Rating of Training Session.....	61
Table B1: SCAAP 2013 Descriptive Statistics of Test Scores by Ethnicity.....	72
Table B2: SCAAP 2013 Ranges of P-values of Test Forms by Ethnic Groups.....	72
Table B3 : 2013 SCAAP Item Analysis Results for Music Form 1 .....	73
Table B4: 2013 SCAAP Item Analysis Results for Music Form 2 .....	75
Table B5: 2013 SCAAP Item Analysis Results for Visual Arts Form 1 .....	77
Table B6: 2013 SCAAP Item Analysis Results for Visual Arts Form 2 .....	79
Table B7: 2013 SCAAP Item P-Values for Music Test Forms by Ethnicity.....	81
Table B8: 2012 SCAAP Item P-Values for Visual Arts Test Forms by Ethnicity .....	83
Table B9: 2013 SCAAP Item P-Values for Music Test Forms by Gender.....	85
Table B10: 2013 SCAAP Item P-Values for Visual Arts Test Forms by Gender.....	87

**LIST OF FIGURES**

*Figure 5.1: P-value Histograms for 2013 Music and Visual Arts Tests* ..... 39

*Figure 7.1. Response to: Rate your level of satisfaction with the support and assistance you received from the SCAAP team during test administration.*..... 65

## I. INTRODUCTION

### SCAAP Purpose and Goals

The South Carolina Arts Assessment Program (SCAAP), established in 2000, is a collaborative effort among the South Carolina State Department of Education (SCDE), the University of South Carolina (USC), and South Carolina arts educators. The purpose of SCAAP is to develop and provide technically sound arts assessments aligned to the *South Carolina Academic Standards for the Visual and Performing Arts*.

In 2013, the SCAAP assessments were used to measure the music and visual arts achievement of fourthgrade students in schools and districts that had received the Distinguished Arts Program (DAP) during the 2012-2013 academic year. This year was the second year that school and districts were permitted to either opt in or opt out of participating in the SCAAP when they applied for DAP funding. The results for all schools are presented in this technical report, and each participating school received an individual school report card. By using SCAAP assessments, arts educators and school district personnel can measure and evaluate their students' arts achievement relative to the *South Carolina Academic Standards for the Visual and Performing Arts* (SCDE, 2003; SCDE, 2010). Examination of the statewide results may be used to help determine the consistency of instructional objectives used in DAPs throughout the state.

### 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 is administered through the Office of Program Evaluation (OPE) within the College of Education at USC. The SCAAP personnel at the OPE for 2013 consisted of a research faculty member (Dr. Tammiee Dickenson) and a senior research associate (Dr. Ashlee Lewis) to lead the project, a project coordinator (Xiaofang Zhang), and a team of graduate research assistants. The research faculty member with expertise in educational research, measurement, and statistics oversaw the project and helped to analyze and monitor student results. In August, 2013, a senior research associate began to lead the team alongside the research faculty member. A research associate served as project coordinator and coordinated project-related tasks. A fiscal coordinator (Becky Derrick) managed the tasks related to collecting grant funds awarded to each of the grantees for the purpose of





participation in SCAAP. The SCAAP team manages all aspects of the SCAAP, including assessment development, coordination of statewide assessment administration, analysis and reporting of student results, and analysis of the psychometric performance of the assessments.

The SCAAP team contracts with E-Business Solutions through USC's University Technology Services to maintain and regularly update the website used to administer the assessment with students. The SCAAP website also serves as a portal through which SCAAP consultants review and score students' responses to the performance tasks.

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

SCAAP Advisory Committees are comprised of arts educators who are acknowledged statewide as leaders in their respective arts areas. Members of the committees have completed the Curriculum Leadership Institute of the Arts (CLIA) and the Arts Assessment Institute (AAI), two of the professional development institutes offered to arts educators in the state by the SCDE. 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, during typical testing cycles, review and refine performance task documents. During Validation Sessions, which are held when performance tasks are administered to students, the Advisory Committee members review and refine performance task documents, and benchmark student performance tasks for subsequent rating. Table 1 provides a list of the core members of the SCAAP Advisory Committees for the 2013 assessment year.

Table 1.1: *List of SCAAP Advisory Committee Members for the 2013 Assessment Year*

Music Advisory Committee	Visual Arts Advisory Committee
Pam Gowan	Connie Boleman
Kathy Clark	Lillie Dunning
Bonnie Pruitt	Maria Robinson

SCAAP History In 2000, the SCAAP project began with a focus on developing and field testing large-scale assessments for elementary school music and visual arts. Advisory Committees, comprised of statewide leaders in music and visual arts education, were formed to determine the test content and format. Advisory Committee members met with the SCAAP personnel from the OPE and SCDE personnel to determine test specifications by selecting the content standards and indicators 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 fourth grade as the elementary grade level to be assessed because the *2003 South Carolina Visual and Performing Arts Curriculum Standards* (SCDE, 2003) used grade level bands; the committee members indicated that teachers would not have had enough instructional time to cover the grades 3-5 standards before testing in the third grade and would not have enough time to use the assessment results to modify instruction if testing occurred in the fifth grade. Therefore, grade 4 was selected as the most appropriate and useful time to measure student achievement.

The assessment formats chosen were multiple-choice and performance tasks. Committee members determined that those two formats were most suited for assessing students' music and visual arts achievement in relation to the *2003 SC Visual and Performing Arts Curriculum Standards* (SCDE, 2003). The Advisory Committee members then created test items and tasks to match the test specifications. In 2002 (Year 1 of test administration), 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 on 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* (Yap, Schneider, Johnson, Mazzie, & Porchae, 2002), submitted to the SCDE.

In 2003 (Year 2), the SCAAP fourth 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 from 40 to 45 items in order to reach adequate reliability. In Year 2, the SCAAP personnel also pilot-tested the web-based music and visual arts assessment prototype. A detailed description of SCAAP activities conducted during Year 2 can be found in the report, *Technical Documentation for the South Carolina Arts Assessment Project (SCAAP) Year 2: Fourth Grade Music and Visual Arts Assessments* (Yap, Johnson, Brown, Moore, & Yoshioka, 2003), submitted to the SCDE. In Year 3, the SCAAP personnel began implementing the web-based SCAAP fourth grade music and visual arts assessments at several schools across the state in conjunction with the DAP grants awarded by the SCDE.

Advisory Committees were formed in Year 3 (2004) to begin developing the entry-level SCAAP dance and theatre assessments. In Year 4 (2005), 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 (2006), 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 (2007) and implemented at several schools across the state in Year 7 (2008). 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.

By Year 7, SCAAP had developed six different assessments, including four entry-level (elementary students) and two intermediate-level (middle school students) 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.

Prior to 2011 (Year 10 of SCAAP), the fourth grade entry level of music and visual arts assessments were constructed based on the *2003 South Carolina Visual and Performing Arts Curriculum Standards* (SCDE, 2003) for grades three through five. With the introduction and implementation of the *2010 South Carolina Academic Standards for the Visual and Performing Arts* (SCDE, 2010), SCAAP personnel collaborated with SCAAP Advisory Committee members to re-align each of the items in the SCAAP item bank (N=169), which is used to construct the multiple-choice test forms to the new academic standards. Items that did not meet the new standards were terminated during the item re-alignment process. After this process was completed, the item bank contained only enough items to develop one multiple choice test form. Therefore, the 2011(Year 10) SCAAP assessment had only one multiple choice test form with 45 items and two performance tasks for both the music and the visual arts.

In the same year (Year 10), SCAAP personnel hosted an item writing retreat with six music and six visual arts expert educators in November of 2011. The purpose of the retreat was to develop assessment items that align with the 2010 academic standards to ensure that an ample pool of items would be available for use with future SCAAP assessments. New test specifications for both music and visual arts assessments were also developed and new multiple-choice items were generated for music and visual arts assessments based on the new academic standards.

In 2012 (Year 11), due to budgetary constraints, only the entry-level music and visual arts multiple-choice assessments were administered. Both assessments were administered during the spring semester, and detailed school-level results were reported to teachers and principals in May 2012.

In the current year (Year 12), both the multiple-choice section and the performance tasks sections were administered to schools who opted to participate in the entry-level assessments. The tests were administered from March 4 to April 19 in 2013.



## Dissemination & Research

As the only web-based and fully-implemented arts assessment in the country with established validity and reliability, SCAAP represents South Carolina's commitment to excellence in arts education. SCAAP researchers have presented information regarding SCAAP assessments and research at scholarly conferences across the country, including the annual conferences of the American Educational Research Association and the American Evaluation Association. 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](mailto:scaap@mailbox.sc.edu) for a complete list of presentations, publications, and research.



## II. TEST FRAMEWORK & CONTENT

### **SCAAP Format and Test Items**

The SCAAP fourth grade entry level test format, developed in Year 1 and modified slightly in Year 2, has traditionally included two 45-item parallel multiple-choice test forms and two performance tasks each for music and visual arts. However, as a result of the item re-alignment to the 2010 academic standards in 2011, Year 10 of SCAAP assessment was administered with only one multiple choice test form with 45 items and two performance tasks for both music and visual arts. After the item retreat that was conducted in late 2011, enough multiple-choice items were created to restore the assessment to its two 45-item parallel test forms in 2012 (Year 11). The SCAAP performance tasks were not administered in Year 11 due to budgetary constraints. In 2013, both multiple-choice and performance tasks were administered to schools who participated in the entry-level assessments.

Each multiple-choice test item used in the SCAAP assessment targets a South Carolina academic standard and has a designated Bloom's taxonomy level. Each multiple-choice item has four response options, and many include multi-media stimuli (i.e., sound files or visual stimuli).

### ***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 for the audio files and jpeg and gif for the image files. Often, the music examples and images used in the SCAAP test items were created by local South Carolina artists and composers.

### ***Item Review***

Many of the multiple-choice items used on the SCAAP assessments from 2004 (Year 3) were created during the initial development phase of the assessment; however, new items have been created over the years to augment the item banks and to replace items terminated due to poor item performance. All items are reviewed and edited yearly by SCAAP 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 academic

standards for the visual and performing arts, and (c) gender and ethnicity bias using Differential Item Functioning (DIF) analysis.

Changes to existing test items are determined by the SCAAP Advisory Committee during item review sessions. Any changes not approved during item review sessions (i.e., changes made by SCAAP personnel) 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 on the SCAAP website. In addition, the SCAAP personnel make sure that all test items adhere to the following item writing guidelines. These item writing guidelines were adapted from those presented in *Put to the Test: Tools and Techniques for Classroom Assessment* (Kuhns, Johnson, Agruso & Monrad, 2001)

### **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 2004 (Year 3), all SCAAP multiple-choice assessments were administered online via the SCAAP website (<https://scaap.ed.sc.edu>). In addition, other aspects of the assessment became web-based, such as performance task rating and monitoring. The SCAAP website is secured by usernames and passwords assigned by the SCAAP personnel who serve as the website administrators. Website administration tasks 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, and (d) monitoring the online remote rating system, among other tasks. Depending on their access level, website users can perform activities such 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. Two new functions were added to the SCAAP website at the beginning of 2011. SCAAP personnel and authorized administrators are now able to maintain and update the item performance history information, and download the item performance history into different formats, such as Excel from the SCAAP website as well.

Prior to the administration of the 2012 assessment, the SCAAP website was re-designed to both change its visual appearance and to add several new features to make the navigation of the site more user-friendly. The SCAAP website uses a SQL\_DB server to store the large amount of data needed to conduct the SCAAP assessments and to accommodate the large volume of concurrent users that occurs during SCAAP test administration and performance task rating. The SCAAP personnel have created the South Carolina Arts Assessment Program Test Administrator Manuals to help facilitate test administrators' navigation of the website for SCAAP assessment usage. The manuals are available to test administrators with individual usernames and passwords only.





## Music Assessment Content & Test Specifications

The SCAAP Music Assessment includes two sections: multiple-choice and performance tasks. The multiple-choice section requires students to demonstrate their knowledge of and skills in (a) music vocabulary, (b) notation, (c) listening to music, (d) evaluation of performance problems, and (e) performance of music skills. Performance tasks require students to demonstrate their singing and rhythm improvisation skills on demand.

### *SC Music Standards and Indicators*

Following are the music standards and indicators selected by the Music Advisory Committee for inclusion in the SCAAP fourth grade music assessment. The music content and achievement standards were selected from *2010 South Carolina Academic Standards for the Visual and Performing Arts* (SCDE, 2010) for grade four.

**Standard 1 (Music Performance):** *The student will sing and perform on instruments a variety of music, alone and with others.*

Indicators:

#### *Singing*

4-1.1: Sing independently on pitch and in rhythm, using appropriate timbre, diction, and posture while maintaining a steady tempo.

4-1.2: Sing expressively, alone or in groups, matching dynamic levels and responding to the cues of a conductor.

4-1.3: Sing, alone and with others, a varied repertoire of music including partner songs, descants, ostinati, and rounds.

**Standard 2 (Creating Music):** *The student will improvise, compose, and arrange music within specified guidelines.*

Indicators:

#### *Improvising*

4-2.1: Improvise short rhythmic and melodic question-and-answer patterns.

4-2.2: Improvise simple rhythmic and melodic ostinati patterns and accompaniments.



4-2.3: Improvise simple rhythmic variations.

4-2.4: Improvise short instrumental pieces using a variety of sound sources.

**Standard 3 (Music Literacy):** *The student will read and notate music.*

Indicators:

4-3.1: Read, write, and perform rhythmic notation incorporating whole, half, quarter, eighth, sixteenth, dotted half notes, and corresponding rests.

4-3.2: Use a system (syllables, numbers, or note names) to read melodic notation and write eight-measure melodic lines in pentatonic and major tonalities.

4-3.3: Identify symbols and terminology for dynamics (*pp*, *mp*, *p*, *mf*, and *f*) tempo

(*allegro*, *presto*, *adagio*, *largo*, *ritardando*, and *accelerando*), and articulation (*staccato* and *legato*).

4-3.4: Use available technology to notate created and dictated measures with standard symbols for meter, rhythm, and pitch.

**Standard 4 (Critical Response to Music):** *The student will listen to, describe, analyze, and evaluate music and music performances.*

Indicators:

*Analyzing*

4-4.1: Identify examples of musical forms (for example, motive, four-bar phrase, echo song, canon, rondo, theme and variation).

4-4.2: Demonstrate perceptual skills by moving to and answering questions about music representing diverse styles.

4-4.3: Use music terminology to describe musical elements (for example, pitch, timbre, notation, meter, chords, tonality).

4-4.4: Identify, describe, and classify by sight and sound a variety of instruments

and instrumental families.

4-4.5: Demonstrate creative movement and emotional response to prominent music characteristics while listening.

*Evaluating*

4-4.6: Devise criteria for evaluating performances and compositions on the basis of music concepts.

4-4.7: Explain his or her preferences for specific musical works and performances.

4-4.8: Describe his or her own performances and those of others and offer constructive suggestions for improvement.

**Standard 5 (History and Culture):** *The student will examine and perform music from a variety of stylistic and historical periods and cultures.*

## Indicators:

4-5.1: Listen to examples of music from various historical periods and world cultures and, from a list, identify the pieces by genres or styles.

4-5.2: Describe ways in which the elements of music (rhythm, tempo, melody, harmony, texture, timbre, articulation, and dynamics) are used in works from various genres and diverse cultures.

4-5.3: Identify uses of music in daily experiences and describe the characteristics that make music suitable for each use.

4-5.4: Discuss and demonstrate audience behaviors for a variety of events (for example, indoor/outdoor concerts, school/community events).

**Standard 6 (Making Connections):** *The student will make connections between music and other arts disciplines, other content areas, and the world.*

## Indicators:

4-6.1: Compare and contrast the roles, careers, and income of musicians in various settings and world cultures.

4-6.2: Identify the role of music in everyday life, celebrations, and other special events.

4-6.3: Integrate music into creative writing, storytelling, poetry, visual arts, and other disciplines (for example, choose a musical instrument to represent the mood of a poem).

4-6.4: Identify connections between mathematics and the rhythmic ideas in music.

### **SCAAP Music Test Specifications**

Table 2.1 presents the test specifications for the SCAAP music assessment. The table presents the overall emphasis of the assessment by standard and how much of the assessment is covered by each assessment format. The test specifications were revised in Year 11 to address the new arts standards. It was applied since 2012 and it will continue to be applied in the future.

Table 2.1: *Table of Specifications for SCAAP Music Assessment*

Standard	Overall Emphasis	Percent covered by Assessment Format	
		Multiple-Choice	Performance Tasks
Standard 1: Performance	25%	-	100%
Standard 2: Creating Music	20%	25%	75%
Standard 3: Music Literacy	25%	100%	-
Standard 4: Critical Response to Music	15%	100%	-
Standard 5: History and Culture	10%	100%	-
Standard 6: Connections	5%	100%	-

### **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).

For Year 10, there was a reduction in the item bank due to the assessment's re-alignment to the 2010 standards. To maintain the integrity of the assessment, the Music Advisory Committee chose to exclusively use items which were closely aligned with the 2010

standards. Therefore, for Year 10, the multiple-choice section of the assessment included one test form.

During Year 11, expert music educators generated 53 items for the music multiple-choice item bank during the three-day SCAAP Item Writing Retreat. This allowed the assessment to return to the use of two parallel 45-item music test forms since 2012. In 2013, five new items were selected for use in the two multiple-choice test forms. Table 2.2 below shows the percentage of items for each multiple-choice form by standard on the 2013 assessment.

Table 2.2: *Percent of Items on Music Multiple-Choice Forms by Standard*

<b>Standard</b>	<b>Percent of Each Multiple-Choice Form</b>
Standard 1: Performance	-
Standard 2: Creating Music	5%
Standard 3: Music Literacy	40%
Standard 4: Critical Response to Music	35%
Standard 5: History and Culture	15%
Standard 6: Connections	5%

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 of the music assessment is 45 points.

### ***Music Performance Task Section: Format & Scoring***

In Year 1, the Music Advisory Committee and SCAAP personnel developed two music performance tasks, which were the same tasks used in Year 12. 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. In 2013, the music prompts and directions for performance tasks were made accessible on a flash drive so that the teacher can play the digital files on a computer or a laptop if they want.

Beginning in Year 3, the SCAAP music performance tasks are scored by trained raters using hierarchical analytic rubrics<sup>1</sup>. After comparing the inter-rater reliability obtained using holistic and analytic rubrics in Year 2, Music Advisory Committee members decided to use hierarchical analytics rubrics because they would (a) allow raters to identify and evaluate the different components of the students' performances separately, and (b) provide participating music teachers with detailed feedback regarding students' strengths and weaknesses. The rubric for Music Performance Task 1 includes three criteria (Tonal, Rhythm, and Vocal Quality) and the rubric for Music Performance Task 2 includes two criteria (Rhythm and Improvisation). Each criterion has 5 score 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 12 music performance tasks are presented in Tables 2.3 to 2.7. Since Year 7 (2008), the music performance tasks rubrics have been provided to test administrators prior to testing.

Table 2.3: *SCAAP 2013 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 or pauses, or stops

<sup>1</sup> 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 2013 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 2013 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 2013 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 2013 Music Task 2 Rubric—Improvisation Criteria

Rating	Improvisation Criteria
4	Improvise a rhythm pattern using complex rhythm patterns such as divisions or elongations
3	Improvise a rhythm pattern using only note values included in the prompt
2	Improvise an 8-beat long rhythm pattern (+ or – one beat)
1	Improvise a rhythm pattern that is 6 beats long or 10 or more beats long
0	Incomplete performance, no improvisation, or no recognizable improvisational intent

Beginning in 2008 (Year 7), raters had the option of choosing augmentation scores (+ and -) for each student performance. These augmentation scores raise the maximum score for each performance criterion in the SCAAP Music Performance Tasks to 4.33. A student's score on each criterion is added together to produce their final score. For Music Task 1, a student can achieve a maximum score of 12.99, and for Music Task 2, the student can achieve a maximum score of 8.66. Table 2.8 presents a summary of the maximum points for each criterion and for each of the SCAAP Music Performance Tasks.

Table 2.8: Maximum Points for Music Assessment Performance Tasks

Performance Task	Criterion	Points per Criterion	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	

### ***SC Visual Arts Standards and Indicators***

Following are the visual arts content and achievement standards selected by the Visual Arts Advisory Committee for inclusion in the SCAAP fourth grade visual arts assessment. The visual arts content and achievement standards were selected from *2010 South Carolina Academic Standards for the Visual and Performing Arts* (SCDE, 2010) for grade four.

**Standard 1 (Creating Works of Visual Art):** The student will demonstrate competence in the use of ideas, materials, techniques, and processes in the creation of works of visual art.

Indicators:

- 4-1.1 Identify the materials, techniques, and processes used in a variety of artworks.
- 4-1.2 Explain the reasons that different elements and principles of design each cause their own distinct response in one who is creating or viewing artworks..
- 4-1.3 Use a variety of media, techniques, and processes to create works of visual art.
- 4-1.4 Select and use the most effective materials, techniques, and processes to communicate his or her ideas, experiences, and stories through works of visual art.
- 4-1.5 Use all art materials and tools in a safe and responsible manner.

**Standard 2 (Using Structures and Functions):** The student will use composition and the elements and principles of design to communicate ideas.

Indicators:

- 4-2.1 Explain the differences in the composition and design of various works of visual art and the ideas they convey.
- 4-2.2 Explain the reasons that different elements and principles of design each cause their own distinct response in one who is creating or viewing artworks.
- 4-2.3 Use visual structures and functions of art to create artworks that communicate ideas.

- 4-2.4 Describe the ways that his or her use of organizational principles and expressive features evoke the ideas he or she intended to convey in a work of visual art.

**Standard 3 (Exploring Content):** The student will examine the content of works of visual art and use elements from them in creating his or her own works.

Indicators:

- 4-3.1 Identify and describe the content in a work of visual art.
- 4-3.2 Select and use subject matter, symbols, ideas, and the elements and principles of design to communicate meaning through his or her art-making.
- 4-3.3 Discuss the ways that choices of subject matter, symbols, and ideas combine to communicate meaning in his or her works of visual art.

**Standard 4 (History and Culture):** The student will understand the visual arts in relation to history and world cultures and the technologies, tools, and materials used by artists.

Indicators:

- 4-4.1 Identify and discuss specific works of visual art as belonging to a particular time, culture, and place.
- 4-4.2 Discuss the qualities of specific works by artists who have had a diverse access to various technologies, tools, and materials.

**Standard 5 (Interpreting Works of Visual Art):** The student will analyze and assess the characteristics and qualities of his or her own works of visual art and those of others.

Indicators:

- 4-5.1 Identify and discuss some of the purposes for the creation of works of visual art.

- 4-5.2 Describe, discuss, and evaluate, both orally and in writing, the different qualities and characteristics of his or her artworks and those of others, including works by South Carolina artists.

### **SCAAP Visual Arts Test Specifications**

Table 2.9 presents the test specifications for the SCAAP visual arts assessments. The table presents the overall emphasis of the assessment by standard and the percentage of the overall assessment that is covered by each assessment format. The test specifications were revised in Year 11 to address the new arts standards. It was applied since 2012, and it will continue to be applied in the future.

Table 2.9: *Table of Specifications for SCAAP Visual Arts Assessment*

Standard	Overall Emphasis	Percent covered by Assessment Format	
		Multiple-Choice	Performance Tasks
Standard 1: Creating Art	25%	40%	60%
Standard 2: Structures and Functions	25%	50%	50%
Standard 3: Exploring Content	10%	100%	-
Standard 4: History and Culture	10%	100%	-
Standard 5: Interpreting Works of Visual Art	25%	25%	75%
Standard 6: Connections	5%	100%	-

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.

For Year 10, there was a reduction in the item bank due to the assessment's re-alignment to the 2010 standards. To maintain the integrity of the assessment, the Visual Arts Advisory Committee chose to exclusively use items that were closely aligned with the 2010 standards.

Therefore, for Year 10, the multiple-choice section of the assessment included one test form of 45 items.

During Year 11, expert visual arts educators generated 57 items for the visual arts multiple-choice item bank during the three-day SCAAP Item Writing Retreat. This allowed the assessment to return to the use of two parallel 45-item visual arts test forms since 2012. In 2013, five new items were selected for use in both of the two visual arts test forms. Table 2.10 below shows the percentage of items for each visual arts multiple-choice test form by standard for the 2013 assessment.

Table 2.10: *Percent of Items on Visual Arts Multiple-Choice Forms by Standard*

<b>Standard</b>	<b>Percent of Each Multiple-Choice Form</b>
Standard 1: Creating Art	25%
Standard 2: Structures and Functions	20%
Standard 3: Exploring Content	15%
Standard 4: History and Culture	15%
Standard 5: Interpreting Works of Visual Art	15%
Standard 6: Connections	10%

Student responses to the 45-item, web-based multiple-choice test 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 of the visual arts assessment 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. SCAAP Visual Arts Performance Task 1 requires students to compare and contrast, in writing, two artworks using a word bank of visual arts terms. The artworks used in this task alternate each year. For Years 1-9, Visual Arts Performance Task 2 was a two-part task. Part one, Task 2a, requires students to complete a drawing based on a given prompt and part two (Task 2b) required students to write a critique of their own drawing using a word bank of visual arts terms. The prompt for Task 2a alternates each year. In 2011 (Year 10), the Visual Arts Advisory Committee made the decision to eliminate Task 2b. This decision was made based on prior psychometric results and on feedback from SCAAP test administrators.



To standardize test administration, each student receives the same performance task booklet, which includes written directions and space to complete the tasks. Test administrators are responsible for making sure that students use # 2 pencils and work independently. Beginning in 2009 (Year 8), each Visual Arts test administrator was provided with a Test Administrator Script to ensure that the delivery of directions was standardized for each student.

Since Year 1, the SCAAP visual arts performance tasks have been scored by trained raters using holistic rubrics. The SCAAP Visual Arts Advisory Committee members decided to use holistic rubrics to describe students' performance levels because the criteria used to evaluate students' performance tasks are dependent on one another and difficult to separate when scoring. Each rubric has 5 levels, ranging from 0 to 4, and each level describes a student's achieved level of proficiency with regard to the specific performance task. The rubrics used to score this year's (Year 12) visual arts performance tasks are presented in Tables 2.11 and 2.12. Beginning in 2009 (Year 8), the rubrics were made available to test administrators prior to administration in both hard copy and on the SCAAP website.

Table 2.11: SCAAP 2013 Visual Arts Task 1 Rubric—Compare/Contrast

0	1	2	3	4
<p>Student uses at most one term correctly when referring to Picture A and B and clearly shows in the context of the writing that he/she understands the art term used.</p> <ul style="list-style-type: none"> <li>Most of the writing is off topic or there is an insufficient amount of original writing to evaluate student's visual arts knowledge.</li> </ul>	<p>Student's writing demonstrates <u>limited understanding</u> of the similarities AND/OR differences between the two artworks.</p> <ul style="list-style-type: none"> <li>Student uses <b>at least two</b> different terms correctly when referring to Picture A and B, and clearly shows in the context of the writing that he/she understands the art terms used.</li> <li>The two terms used may not include any specific references or explanations.</li> <li>The general references must demonstrate that the student understands the art term.</li> </ul>	<p><b>Some</b> of the student's writing demonstrates a <u>clear understanding</u> of the similarities AND/OR differences between the two artworks.</p> <ul style="list-style-type: none"> <li>Student uses <b>at least three</b> different terms correctly when referring to Picture A and B, and clearly shows in the context of the writing that he/she understands the art terms used.</li> <li><b>All three</b> terms used are <b>general OR specific</b> references.</li> <li>The terms <b>may or may not</b> include <b>explanations</b>.</li> </ul>	<p><b>Most</b> of the student's writing demonstrates a <u>clear understanding</u> of the similarities AND differences between the two artworks.</p> <ul style="list-style-type: none"> <li>Student uses <b>at least four</b> different terms correctly when referring to Picture A and B, and clearly shows in the context of the writing that he/she understands the art terms used.</li> <li><b>At least two</b> of the four terms used are <b>specific references</b>.</li> <li><b>At least one</b> of the four of the terms used is an <b>explanation</b>.</li> </ul>	<p><b>Most</b> of the students' writing demonstrates a <u>clear understanding</u> of the similarities AND differences between the two artworks.</p> <ul style="list-style-type: none"> <li>Student uses <b>at least four</b> different terms correctly when referring to Picture A and B, and clearly shows in the context of the writing that he/she understands the art terms used.</li> <li><b>Three</b> of the four terms used are <b>specific references</b>.</li> <li><b>At least two</b> of the four terms used are <b>explanations</b>.</li> <li><b>Only one</b> additional term may be used <b>incorrectly</b>.</li> </ul>

Table 2.12: SCAAP 2013 Visual Arts Task 2 Rubric—Drawing

0	1	2	3	4
<p>The drawing:</p> <ul style="list-style-type: none"> <li>Does NOT address assigned topic,</li> </ul>	<p>The drawing:</p> <ul style="list-style-type: none"> <li>Includes the assigned topic and an attempt was made to fill the space.</li> </ul>	<p>The drawing:</p> <ul style="list-style-type: none"> <li>Includes the assigned topic and fills most of the space.</li> </ul>	<p>The drawing:</p> <ul style="list-style-type: none"> <li>Includes the assigned topic and fills the space but <b>may or may not be unified.</b></li> </ul>	<p>The drawing:</p> <ul style="list-style-type: none"> <li>Includes the assigned topic and fills the space in a unified composition.</li> </ul>
<ul style="list-style-type: none"> <li>Does NOT attempt to fill the space, or drawings are unrecognizable.</li> </ul>	<ul style="list-style-type: none"> <li>Shows NO attempt to include a background and foreground and to create a sense of depth in the composition.</li> </ul>	<ul style="list-style-type: none"> <li>Indicates attempts to include an environment and create depth, but there is <b>NOT a clear distinction</b> between background and foreground.</li> </ul>	<ul style="list-style-type: none"> <li>Includes an environment with a <b>clear distinction</b> between the background and foreground that <b>may or may not</b> create a sense of depth in the whole composition.</li> </ul>	<ul style="list-style-type: none"> <li>Includes an environment with a clear distinction between the background and foreground creating a sense of depth in the whole composition.</li> </ul>
	<ul style="list-style-type: none"> <li>Includes very limited details, texture, or pattern.</li> </ul>	<ul style="list-style-type: none"> <li>Includes few details, texture, or patterns.</li> </ul>	<ul style="list-style-type: none"> <li>Includes some details, texture, or patterns.</li> </ul>	<ul style="list-style-type: none"> <li>Includes many details. Texture and patterns are used to enhance the picture.</li> </ul>



Since 2008 (Year 7), raters have had the option of using augmentation scores (+ and -) when scoring students' performances. Therefore, the maximum scores for Visual Arts Tasks are 4.33. 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*

<b>Performance Task</b>	Description	<b>Points per Criteria</b>	<b>Points per Task</b>
Task 1	Compare/Contrast	4.33	4.33
Task 2	Drawing	4.33	4.33

### III. TEST ADMINISTRATION

#### Participants

In 2013 (Year 12), 18 schools and six school districts received DAP grants from the SCDE and opted to participate in SCAAP. Of the six participating school districts that received district-level DAP grants, five districts registered three elementary schools to participate in the SCAAP fourth grade assessments. One small district registered the only elementary school in the district to participate in SCAAP. The six school districts with district-level DAP grants registered a total of 16 schools for Year 12 SCAAP testing. A total of 34 DAP elementary schools from around the state participated in the Year 12 SCAAP assessments. From the participating schools, 2,599 students completed the music assessment, and 2,763 students completed 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 12, test administrators who were new to the assessment were required to complete the SCAAP Test Administrator Training Session. The training was administered using an interactive webinar that included a live, on-camera trainer, a PowerPoint presentation describing various elements of test administration, and a live demonstration on the use of the SCAAP website. In addition to being offered as a live session in which test administrators could interact through a chat function, the session was recorded and the link to the recorded training was emailed to all the test administrators. The link to the training session is available on the SCAAP website and may only be viewed by test administrators with assigned logins. This procedure allowed test administrators to watch the training video according to their schedules.

The purpose of the training was to familiarize the test administrators with the SCAAP administration requirements and procedures including managing registered students on the SCAAP website, administering the web-based multiple-choice test forms and music performance tasks. All information regarding the SCAAP assessment procedures was also documented in the South Carolina Arts Assessment Program Test Administration Manual for 2013. The manual was emailed to test administrators along with their login information for the website and their list of students' login information. In addition, electronic copies of the manual were available on the SCAAP website to test administrators with individual logins. 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 fourth grade music and visual arts assessments, provided that testing began no earlier than March 4, 2013 and was completed no later than April 19, 2013, a total of six weeks. This window was selected based on the Measures of Academic Progress (MAP) testing schedules that were provided by participating schools at the beginning of the school year. At least 100 fourth grade students in each school were required to complete the multiple-choice section of the music and visual arts assessments. In schools with less than 100 fourth graders, all fourth grade students completed the multiple-choice section of the assessment. Additionally, 40 students from each participating school were required to complete the music and visual arts performance tasks. Those 40 students needed to have also completed the multiple-choice section of the assessment.

### ***Multiple-Choice Administration: Music & Visual Arts***

Although the SCAAP assessments are not timed, approximately one hour was allocated for completing each subject area presented in the web-based multiple-choice section, which had 45 items for each subject area. The music multiple-choice test form was divided into two sections: "Understanding Music" and "Listening to Music"; the visual arts multiple-choice test form was not divided into sections. Students worked individually on computers and answered questions at their own pace. Each student wore headphones for the listening portion of the test, 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 needed.

### ***Performance Task Administration: Music***

The SCAAP music performance tasks were administered individually to students. The music test administrators administer the tasks following a standardized procedure. The test administrators (a) play the music prompts and directions for each student, (b) digitally record that student's performance using a microphone provided by SCAAP, and (c) save that student's performance on a provided flash drive using a file naming convention. Test administrators return the flash drives to the SCAAP personnel using a business reply envelope provided by SCAAP. After the files are returned, the SCAAP personnel then employ a mass uploading procedure to place 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 2013* and are presented in detail in the test administrator training video. In addition, the SCAAP personnel provided technical support during regular school hours.

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

### ***Performance Task Administration: Visual Arts***

The SCAAP visual arts performance tasks were administered to groups of students. Each student is provided with a copy of the SCAAP Visual Arts Performance Task booklet. In the SCAAP test administrator training manual video, the visual arts test administrators are instructed to administer the tasks following a standardized procedure. The test administrator manual includes a detailed preparation checklist and the preparation procedures for the performance tasks. Test administrators are asked to assist their students in writing their usernames and passwords on each page of the performance task booklet and to make sure that students use only a # 2 pencil to complete their drawing. A script was provided to visual arts teachers to help them lead their students through the testing process for the performance tasks in a standardized way. For the visual arts performance tasks, a maximum of 60 minutes was allotted to complete both tasks—approximately 30 minutes per task (i.e., Task 1, Task 2).

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 tasks are included in the *South Carolina Arts Assessment Program Test Administration Manual 2013*.

## IV. SCORING

### **Multiple-choice section**

For the web-based multiple-choice test forms, student responses were stored on the SCAAP website. The SCAAP personnel visually reviewed the results for data cleaning, and the students' answers were scored using an answer key stored in the item database. The SCAAP personnel also downloaded the database from the website for further statistical analyses using classical item theory, including the calculation of reliability indices for the test forms and item-level analyses including item difficulty, discrimination indices, and differential item functioning analysis.

### **Performance Task Section**

#### ***Benchmarking Music and Visual Arts Performance Tasks***

Each year prior to performance task rating, Music and Visual Arts Advisory committees meet with the SCAAP personnel to review and finalize the performance task rubrics from the previous year and to benchmark student performances. In 2013 (Year 12), about 100 benchmarked performances were identified for music performance tasks and 155 for visual arts performance tasks. All of the music benchmarks were recycled from previous years, and they included benchmarks from every year from 2008 to 2011. And about one third of the visual arts benchmarked tasks were newly selected and validated this year (2013), and the remaining benchmarked tasks were selected from previous years tasks (2010 and 2011).

#### ***Web-based Rating System and Procedure***

This year, three experienced raters for each arts area were invited to be SCAAP raters. Each of these raters have been rating SCAAP performance tasks for multiple years and are very experienced with the rating procedures and the web-based rating system. Therefore, the training materials and procedures were uploaded to the SCAAP website. Training materials included rater manuals, scoring rubrics, anchor tasks with score justifications, and a practice test. Raters reviewed training materials to prepare themselves for the rating process. The rater manual provides technical information regarding the web-based system and the rating procedure. The raters were given three weeks to complete the training themselves as well as the rating tasks. All raters completed the rating of all performance tasks within three weeks.

The student performances that were 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. The raters were required to read the rater manual and review the anchor tasks first, and then, they participated in the practice test training. During the practice test training, the raters practiced by completing a practice test which was made in an interactive PowerPoint format. 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 he/she achieved a passing score.

After raters passed the qualifying test, they were permitted to rate student performances remotely via the website according to their own schedule. Based on the number of tasks in Year 12, 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 625 student responses per performance task and approximately 1,250 student responses in total. Raters were randomly assigned to rater groups.

The SCAAP personnel monitored 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% correct. 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 (i.e., nonadjacent scores). A rater is not permitted to continue rating until s/he passes the refresher test.

The SCAAP score resolution method for nonadjacent scores utilizes expert raters, which is a widely accepted rating practice (Johnson, Penny, Fisher, & Kuhs, 2000; Johnson, Penny, Fisher, & Kuhs, 2003; Johnson, Penny, Gordon, 2001). For Year 12, raters were able to use augmentation scores (+ or -) when grading the responses. Previous studies have indicated that score augmentation tends to improve inter-rater reliability (Penny, Johnson, and Gordon, 2000). This

year, about 50% of the music and visual arts performance tasks were double-rated, (i.e., each of those student performances was scored by two raters). The remaining student responses were each scored by a single rater.

## V. ANALYSIS & RESULTS

### Multiple-Choice Section

#### *Analysis*

The results presented in this section include reliability and descriptive statistics for both 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) methods to equate the test forms based on the students' abilities. The SCAAP personnel used BILOG software to perform the computations.

#### *Reliability*

All multiple-choice items were binary (i.e., scored 0 or 1). Classical reliability indices were computed to obtain reliability for each test form. 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 ( $\theta$ ) for both forms. The formula for the empirical reliability is

$\frac{\text{Variance}(\theta)}{\text{Variance}(\theta) + \text{Variance}(\text{Error})}$ . Table 5.1 contains the classical reliability indices for each test form and the empirical reliability for the equated scores.



Table 5.1: *SCAAP 2013 Reliability Indices for Music and Visual Arts Multiple-Choice Items*

<b>Test</b>	<b>Empirical Reliability</b>	<b>Form</b>	<b>Number of Items</b>	<b>Cronbach's Alpha</b>	<b>Corrected Split-Half</b>
Music	0.83	1	45	0.79	0.79
		2	45	0.82	0.82
Visual Arts	0.86	1	45	0.85	0.86
		2	45	0.86	0.85

The reliability estimates for the 2013 SCAAP assessments are within a range that is appropriate for making medium-stakes decisions at the school level. For the 2013 SCAAP test forms, the classical reliability indices ranged from .79 to .86. It is recommended that a minimum reliability index of .70 be used for low-stakes decision making and for research and evaluation purposes (Herman, Ashbacher, & Winters, 1992), and the 2013 SCAAP results are well within that range.

While SCAAP does not currently provide individual student results, the reliability indices are approaching the range that is recommended for individual student reporting. There are several perspectives on the minimum reliability index needed to report individual student results. Philip (2000) reports that a minimum reliability index of .85 is needed for a test form to be used for making high stakes decisions about individual students; however, numerous other sources have reported that in applied settings with high stakes, the minimum reliability index should be .90 (Herman, Aschbacher & Winters, 1992; Kaplan & Saccuzzo, 1982; Nunally, 1978). According to Thorndike, Cunningham, Thorndike, and Hagen (1991), a reliability of .80 for individual scores will produce a stable mean for a group of at least 25 people, and Hill (2002) found that increasing the group size will also increase the reliability estimates. Therefore, the reporting of individual student scores may be possible in future applications of the SCAAP, but will need further consideration, in particular for the music assessment.

The music multiple-choice test was divided into two sections, "Understanding Music" and "Listening to Music." The Understanding Music 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 reliability estimates for the two sections of the Music multiple-choice are not recommended for comparison; each of the two sections has a different number of items, which can impact the reliability estimates.

Table 5.2: SCAAP 2013 Reliability Indices for Music Multiple-Choice of “Understanding and Listening” sections

	Form 1		Form 2	
	Reliability	Number of Items	Reliability	Number of Items
Understanding	0.71	30	0.76	31
Listening	0.64	15	0.67	14

Each SCAAP multiple-choice item is designed to address a particular content standard based on the 2010 South Carolina Academic Standards for the Visual and Performing Arts (SCDE, 2010). The SCAAP researchers computed reliability indices for each standard. Those indices for each music standard are presented in Table 5.3; indices for visual arts standards are in Table 5.4.

Table 5.3: SCAAP 2013 Reliability Indices for Music Multiple-Choice Assessment by SC Standard

	Form 1		Form 2	
	Reliability	Number of Items	Reliability	Number of Items
Standard 2	-. <sup>2</sup>	2	-. <sup>3</sup>	2
Standard 3	0.56	18	0.66	18
Standard 4	0.65	16	0.65	16
Standard 5	0.44	7	0.47	7
Standard 6	0.18	2	0.16	2

<sup>2</sup> There were only two items for this standard. The reliability was too low to report. This is likely due to the lack of variability which might cause the very low reliability estimate for this standard.

<sup>3</sup> There were only two items in this standard. The reliability was too low to report. This is likely due to the lack of variability which might cause the very low reliability estimate for this standard.

Table 5.4: *SCAAP 2013 Reliability Indices for Visual Arts Multiple-Choice Assessment by SC Standard*

	Form 1		Form 2	
	Reliability	Number of Items	Reliability	Number of Items
Standard 1	0.65	11	0.59	11
Standard 2	0.44	9	0.50	9
Standard 3	0.52	8	0.35	6
Standard 4	0.47	6	0.55	8
Standard 5	0.45	7	0.56	7
Standard 6	0.25	4	0.35	4

Considering that, generally, “short tests or subtests simply give less reliable scores than they would if they were composed of more items” (Mehrens, & Lehmann, 1991, p. 258), the relatively low reliability indices reported 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. In SCAAP 2013, the Music Advisory Committee decided to include two new Standard 2 items in the multiple-choice assessment. The reliability of Standard 2 in both forms were low due to the small number of items. 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 remained stable since 2010 when items were re-aligned with the new academic standards.

### ***Descriptive Statistics***

The mean scores were 21.5 and 22.9 for the Music Forms 1 and 2, respectively; for Visual Arts Forms 1 and 2, the mean scores were 25.1 and 24.4 (see Table 5.5). These means indicate that students correctly answered approximately half of the items on both the music and visual arts test forms on average. For a test item with four answer options, a student with no knowledge will have a one in four chance of getting the correct answer by pure guessing. Therefore, for a test of 45 items, we estimated that a student might score approximately 11 points if they guessed for all items. The achieved means for the multiple-choice assessments indicate that

students' results are based not on chance, but on their understanding of the music and visual arts concepts covered in the assessment.

Table 5.5: SCAAP 2013 Descriptive Statistics for Music and Visual Arts Test Forms

	Form	<i>M</i>	<i>SD</i>	<i>N</i>	Number of Items
Music	1	21.5	6.6	1305	45
	2	22.9	7.1	1294	45
Visual Arts	1	25.1	7.7	1361	45
	2	24.4	7.9	1402	45

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

### **IRT Model Fit**

Using Item Response Theory (IRT) methods to analyze large-scale assessment data offers several advantages over classical test theory methods. The main advantage of using IRT for the SCAAP assessments is that it allows the equating of test forms based on latent variables across years and across test forms. That is, student scores can be analyzed to ensure that they are not affected by the test's difficulty level, as the test difficulty may not be the same across years. This is accomplished by using similarly functioning items across different test forms and different years. Another advantage is the ability to use individual item information to guide the item review process.

To fully realize the benefits of IRT methods, the IRT model (i.e., one-, two-, or three-parameter logistic models<sup>4</sup>) used for parameter estimation must fit the data. To examine model fit and to determine the most appropriate model, the analysis of the residuals with the aid of graphs is the method most commonly proposed by researchers (Hambleton, Swaminathan, & Rogers, 1991; Hambleton & Swaminathan, 1985). In the IRT analysis for this year (Year 12) and for previous years, the SCAAP personnel used the two-parameter logistic model because it provided a very good fit for the SCAAP data. A detailed description of the model fit process used by the SCAAP

<sup>4</sup>In the one-parameter logistic model, the probability for an examinee to correctly answer an item *i* is the product of the examinee's ability ( $\theta$ ) and the item difficulty ( $b_i$ ). The two-parameter logistic model involves two parameters characterizing item *i*, item difficulty ( $b_i$ ) and item discrimination ( $a_i$ ). The three-parameter logistic model is based on the two-parameter model by adding a third parameter, guessing, denoted here as  $c_i$  (Hambleton, Swaminathan, & Rogers, 1991).

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* (Yap, Moore & Peng, 2004), submitted to the SCDE.

### ***Equating Test Forms***

For each arts area, the SCAAP multiple-choice section consisted of two parallel test forms of 45 items on each, with 24 linking items between the two test forms; test forms were constructed to be parallel based on the appropriate Table of Specifications (music or visual arts). As students logged into the SCAAP website, they were automatically randomly assigned to one of the two test forms. The concurrent calibration method was then used to compute equated test scores for individual schools. The concurrent calibration method was chosen because this method yields more stable equated scores than other methods such as linear and equipercentile equating (Petersen, Cook, & Stocking, 1983; Hills, Subhijah, & Hirsch, 1988). This method requires the creation of a combined data set that includes all students and all items from both test forms. In this combined data set, when an item was included in the test form the student took, but the student did not provide an answer to the item, the response was coded as 8 (no response). The items not included in the test form the student took were coded as 9 (missing). For example, if Student A completed music test form 1, each item in form 1 was coded as 1 (correctly answered), 0 (incorrectly answered), or 8 (not answered), while all items used in form 2 would be coded as 9 (missing) 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. BILOG default settings were used for the item parameters estimation procedure and the specification of prior distributions on the item parameters (see Appendix A for a sample of the BILOG syntax used in this analysis).

For this year, SCAAP school-level results were computed by equating across test forms. Because individual test forms were comprised of 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, classical 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, DIF indices based on gender and ethnicity were computed for each item. The individual item analysis results for each item in each test are presented in Appendix B; see Tables B3 and B4 for music and Tables B5 and B6 for visual arts.

Histograms of the  $p$ -values, or proportion of correct responses, for the items on each of the test forms were constructed to investigate the distribution of  $p$ -values; these 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. In addition, the most common  $p$ -value for the music test is around .40, whereas the most common  $p$ -value for the visual arts test is around .70, which suggests that on average, this year's music test was more difficult than the visual arts test. Therefore, we would expect that the means for each school's visual arts results would be higher than for the music results.

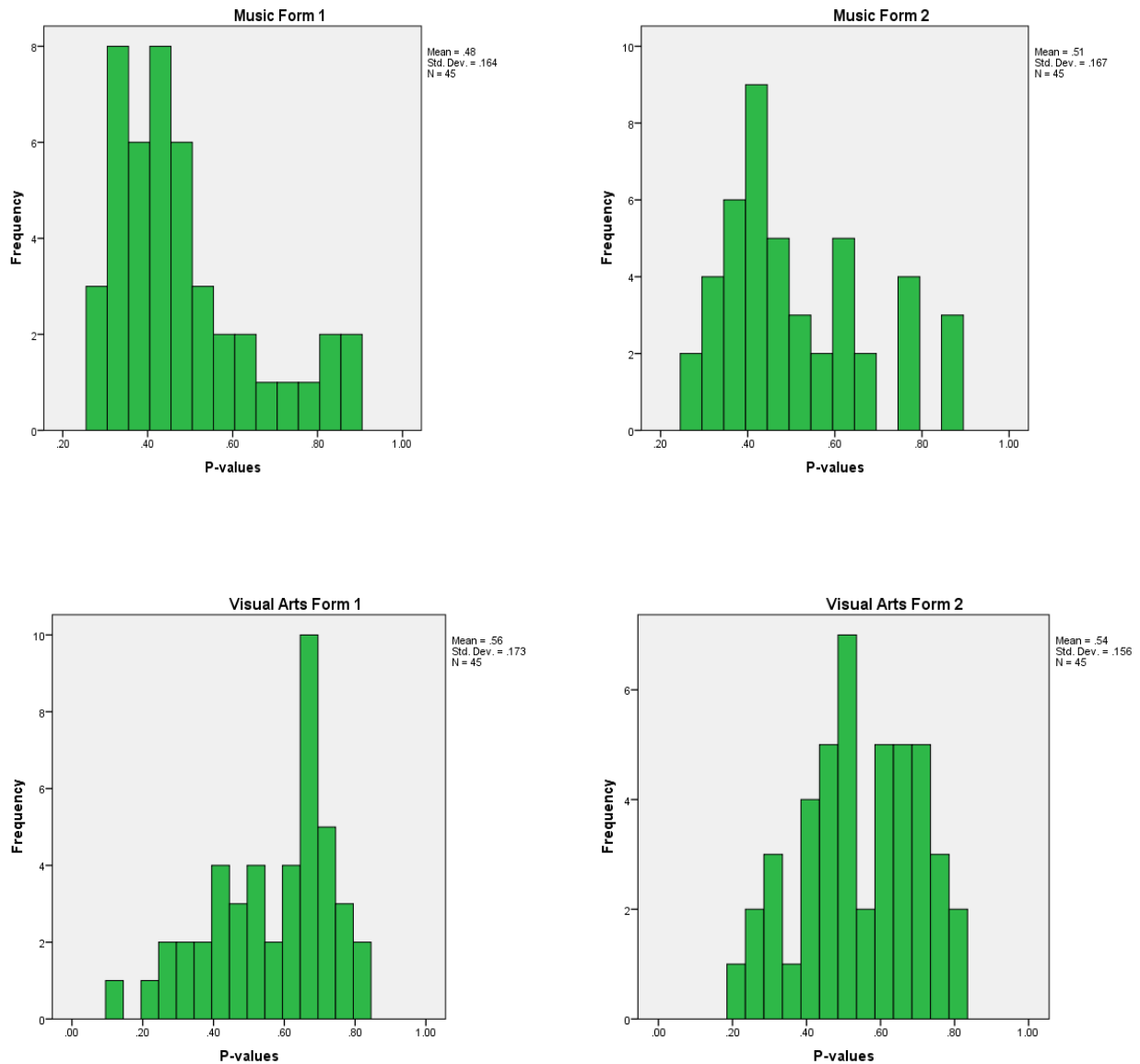


Figure 5.1:  $P$ -value Histograms for 2013 Music and Visual Arts Tests

The ranges of item  $p$ -values and item discrimination indices for each test are presented in Table 5.6. Individual item  $p$ -values ranged from .12 to .88 (see Table 5.6), and the average  $p$ -values for all the forms ranged from .48 to .56 (see Figure 5.1). This year, the average  $p$ -values for the two music forms were lower than the visual arts, and the ranges of  $p$ -values for two areas were similar, indicating that both music and visual arts forms have variability in item difficulty, and have some very difficult items with low  $p$ -values. The item  $p$ -values for music and for visual arts disaggregated by ethnicity are presented in Tables B7 and B8 of Appendix B, and the  $p$ -values for music and visual arts by gender are presented in Tables B9 and B10 of Appendix B.

Table 5.6: SCAAP 2013 Ranges of  $P$ -values and Discrimination Indices of Test Forms

Content Area	Form	P-value Ranges	Discrimination Indices Ranges
Music	1	0.28 – 0.86	0.10 – 0.46
	2	0.27 – 0.88	0.15 – 0.49
Visual Arts	1	0.12 – 0.80	0.06 – 0.53
	2	0.21 – 0.81	0.22 – 0.51

Individual item discrimination indices for all test items ranged from .06 to .53 (see Table 5.6). The item discrimination index is a type of correlation in which each item is correlated with the total test score. A high item discrimination index indicates 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 are considered to be good items, (b) between .30 and .39 are considered reasonably good items, (c) between .20 and .29 are considered marginal items that need minor revision, and (d) equal or lower than .19 are 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 (.19 or less)

is listed in Table 5.7. Two items were common on both forms for the music assessment. Those items with low discrimination indices will be presented to the Advisory Committees at the Item Review Sessions. Based on the committee members' recommendations, those items may be altered or terminated in preparation for Year 13 (2014) testing.

Table 5.7: SCAAP 2013 Number of Items with Low and Negative Discrimination Indices

<b>Content Area</b>	<b>Form</b>	<b>Low Discrimination Indices (<math>0.00 \leq D \leq 0.19</math>)</b>	<b>Negative Discrimination Indices</b>
Music	1	7	0
	2	5	0
Visual Arts	1	2	0
	2	0	0

To investigate whether the student results differ by gender or by ethnic group, the SCAAP personnel computed descriptive statistics for both tests 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 and Hispanic students on both music and visual arts assessments. The ranges of  $p$ -values for each test form are presented in Table 5.10 by gender and ethnicity.  $P$ -value ranges were similar for gender groups, but a slight difference existed among the ethnicity groups, with white students scoring a little higher than their African American and Hispanic counterparts. The  $p$ -values for other ethnicity group are presented in Table B1 and B2 in Appendix B.



Table 5.8: SCAAP 2013 Descriptive Statistics of Test Scores by Gender

Content Area	Form	Gender					
		Female			Male		
		M	SD	n	M	SD	n
Music	1	21.9	6.6	671	21.7	6.5	634
	2	23.6	7.3	652	22.2	6.9	642
Visual Arts	1	26.1	7.6	738	23.9	7.6	623
	2	25.1	7.7	679	23.6	8.0	723

Table 5.9: SCAAP 2013 Descriptive Statistics of Test Scores by Ethnicity

Content Area	Form	Ethnicity								
		African American			Hispanic			White		
		M	SD	n	M	SD	n	M	SD	n
Music	1	19.2	5.9	618	20.3	5.3	90	24.1	6.4	554
	2	20.6	6.6	589	21.4	6.1	81	25.5	6.9	576
Visual Arts	1	21.8	6.7	653	24.7	5.9	77	28.7	7.2	583
	2	21.1	7.2	673	23.9	6.3	94	28	7.3	581

Table 5.10: SCAAP 2013 Ranges of P-values of Test Forms by Gender and Ethnic Groups

Content Area	Form	P-value Ranges				
		Female	Male	African American	Hispanic	White
Music	1	0.24-0.90	0.28-0.85	0.22-0.78	0.22-0.91	0.32-0.94
	2	0.28-0.91	0.24-0.86	0.21-0.83	0.23-0.89	0.31-0.94
Visual Arts	1	0.12-0.84	0.13-0.79	0.15-0.75	0.08-0.88	0.11-0.86
	2	0.23-0.84	0.19-0.81	0.18-0.75	0.20-0.88	0.24-0.88

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

The DIF analysis used is based on the Mantel-Haenszel (MH) procedure, which is included as a routine procedure in the SAS statistical software package. The routine used 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 former South Carolina statewide assessment, the Palmetto Achievement Challenge Test (PACT). Specifically, the MH procedure first categorizes 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 the SCAAP gender DIF analysis, the focal group refers to females and the reference group is males. For the SCAAP ethnicity DIF analysis, African American students are the focal group

and white students are the reference group. The ethnicity analysis was isolated to African American and white students due to relatively small population of other ethnic subgroups in the South Carolina student population.

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

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

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

In SCAAP 2013 (Year 12), all items were classified as either “A” or “B” items when DIF was examined in terms of gender groups, which means the probability of answering an item correctly is very close for male and female students when their proficiency on the measured construct is similar. For DIF in terms of ethnicity groups, most items were “A” and “B” items. Two music items were classified as “C” items, and these two music items require further review. A summary of the DIF classifications for music and visual arts items is presented in Table 5.11.

Typically, “C” items are reviewed for possible sources of bias. Both of the “C” music items tested this year were tested in the previous years. One of them was classified as a “C” item last

year but it was either “A” or “B” in the previous years, while the other one was classified as “A” in previous administrations. SCAAP personnel noted that one of the “C” items, which significantly favors African American students on music forms 2 (Item ID 372), is closely related to African American culture. This item will be reviewed by the Advisory Committee again before Year 13 testing to decide whether to terminate the item or to revise it. The Advisory Committees will be consulted regarding whether to include the other item in the upcoming year of testing as they are or whether those items should be revised. Individual ETS deltas for gender and ethnicity and their respective p-values for each item in each test form are presented in Tables B3 to B10 in Appendix B.

Table 5.11: SCAAP 2013 Summary of DIF Classification for Multiple-Choice Test Forms

	Forms	Reference Group	Focal Group	Total N of Items	DIF Classification		
					A	B	C
Music	1	Male	Female	45	41	4	0
		White	African-American	45	39	6	0
	2	Male	Female	45	43	2	0
		White	African-American	45	36	7	2
Visual Arts	1	Male	Female	45	44	1	0
		White	African-American	45	42	3	0
	2	Male	Female	45	44	1	0
		White	African-American	45	43	2	0
All Forms		Male	Female	180	172	8	0
		White	African-American	180	160	18	2

### **Performance Task Section**

Students' music performance tasks were rated by four raters (including an expert rater) using the SCAAP Music Performance Task rubrics; about 50% of the tasks were each rated by a pair of raters and the rest were rated by a single rater. Each of the four raters scored approximately 600 responses for each performance task in music. In cases where raters disagreed by more than one score point (i.e., "nonadjacent agreement"), responses were scored again by an expert rater. Students' visual arts performance tasks were rated using SCAAP Visual Arts Performance Task rubrics. Four raters were employed (including an expert rater), and each of them rated approximately 1,260 performance task responses. Approximately 50% of the visual arts responses were scored by a pair of raters. For the double-rated responses, scores for each student's performance task were calculated by averaging the scores of the two raters. In cases of nonadjacent agreement, an expert rater's score was used instead. In 2012 (Year 11), augmentation scores were used by raters by adding a "+" or "-" to their scores. Augmentation of scores increases or decreases a numerical score by 0.33.

### ***Inter-Rater Reliability of Performance Tasks***

Generalizability (G) theory was used to estimate inter-rater reliability for each performance task. When using G theory, a coefficient is obtained that reflects "the accuracy of generalizing from a person's observed score on a test or other measure (e.g., behavior observation, opinion survey) to the average score the person would have received under all possible conditions that the test user would be equally willing to accept" (Shavelson & Webb, 1991, p.1). The computer program GENOVA (Crick & Brenna, 1983) was used to estimate the G-coefficients and indices of dependability to obtain information regarding sources of variation in the measurement. The index of dependability for each performance task takes into account a shift in means due to rater effects. Although multiple rater-groups were involved in scoring the performance tasks, one-facet design, P x R was used as recommended by Lee, Kantor, & Mollaun (2001). They asserted 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 and the visual arts performance tasks. As discussed earlier, the music performance tasks employ analytic rubrics based on multiple criteria, while the visual arts performance tasks are scored based on holistic rubrics. 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 two raters were computed.

Table 5.12: *SCAAP 2013 Inter-Rater Reliability of Music Performance Tasks Using Analytic Rubrics*

<b>Performance Task</b>	<b>Criteria</b>	<b>Generalizability Coefficient</b>	<b>Index of Dependability</b>
1 (Singing)	Tonal	0.94	0.94
	Rhythm	0.84	0.83
	Vocal	0.87	0.87
2 (Improvisation)	Rhythm	0.87	0.87
	Improvisation	0.79	0.79

Table 5.13: *SCAAP 2013 Inter-Rater Reliability of Visual Arts Performance Tasks Using Holistic Rubrics*

<b>Performance Task</b>	<b>Generalizability Coefficient</b>	<b>Index of Dependability</b>
1 (Compare and Contrast)	0.88	0.88
2 (Drawing)	0.74	0.74

In 2013, the generalizability coefficients and dependability indices are the same for each of the visual arts performance tasks, indicating no rater leniency/severity effect for the tasks. The generalizability coefficients and indices of dependability are the same for Music Performance Task 1 and Task 2, except for the rhythm criterion in Task 1, which has a very small difference of .01. The results suggest that the two raters had a very close understanding of the rubrics or that they interpret of students' performances in the same way. In general, the closeness of the generalizability coefficients and the dependability indices suggests little to no rater effect for SCAAP tasks. Notably, the generalizability coefficients and dependability indices for the rhythm criterion for Music Task 1 were slightly higher than the previous year, but the generalizability coefficients and dependability indices for the improvisation criterion for Task 2 were slightly lower than for Year 10. These changes indicate that students' performance improved in their

performance of the rhythm aspect of Task 1, but their performance declined in the improvisation aspect of Task 2. Both the generalizability coefficients and dependability indices for Visual Arts Task 1 were the same compared to last year's results, but those for Task 2 were lower. Such changes might be attributed to differences in the visual arts prompts used in odd-numbered and even-numbered years.

### ***Descriptive Statistics***

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

The music analytic rubric scores indicate specific information regarding students' achieved skill level. A score derived from an analytic rubric indicates that the student has achieved the descriptive information contained in the achieved level and in all preceding levels. See Chapter 2 for the analytic rubrics used to score the SCAAP Music Performance Tasks. For Music Task 1: Singing, the mean score for the tonal criterion was 2.63, 3.75 for the rhythm criterion and 2.89 for the vocal criterion. All of these scores were similar results of the previous year (2010-2011). Following is an interpretation summarizing the mean scores for each Task 1 criterion.

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

For Music Task 2: Rhythm Improvisation, the mean score for the rhythm criterion was 3.28 and 2.60 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: SCAAP 2013 Descriptive Statistics for Music Performance Tasks Using Analytic Rubrics

<b>Task</b>	<b>Dimensions</b>	<b><i>M</i></b>	<b><i>SD</i></b>	<b><i>N</i></b>
Singing	Tonal	2.63	1.37	1139
	Rhythm	3.75	0.60	1139
	Vocal	2.89	0.90	1137
	Task Total	9.28	2.48	1137
Improvisation	Rhythm	3.28	0.99	1018
	Improvisation	2.60	1.23	1018
	Task Total	5.89	1.75	1080
<b>Total Score</b>		<b>15.20</b>	<b>3.34</b>	<b>1003<sup>5</sup></b>

The descriptive statistics for the Visual Arts Performance Tasks are presented in Table 5.15. Students earned an average of 4.36 points out of 8.66 possible points for the combined visual arts performance tasks. See Chapter 2 for copies of the holistic rubrics used to score the SCAAP visual arts performance tasks. For Visual Arts Task 1 (Compare and Contrast), the mean score was 2.21 with a standard deviation of 1.11. 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 2 (Drawing) was 2.14 with a standard deviation of 0.79. Following is an interpretation of the mean score for Visual Arts Task 2.

<sup>5</sup> If students did not complete both music performance tasks, their scores were not included in the computation of the total score.

*In general, students completed a drawing that addressed the assigned topic and filled 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 clear distinction between the foreground and the background. The average student included some detail, texture, and pattern to enhance their drawing.*

Table 5.15: SCAAP 2013 Descriptive Statistics for Visual Arts Performance Tasks Using Holistic Rubrics

<b>Performance Task</b>	<b>M</b>	<b>SD</b>	<b>N</b>
1	2.21	1.11	1184
2	2.14	0.79	1184
<b>Total Score</b>	<b>4.36</b>	<b>1.48</b>	<b>1184</b>

**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 were .71 for Task 1 and .68 for Task 2, and the *p*-values for visual arts performance tasks were .51 and .49. The *p*-values for the music and visual arts performance tasks are presented in Table 5.16.

Table 5.16: SCAAP 2013 Task Difficulty Indices and Discrimination Indices for Music and Visual Arts Performance Tasks

	<b>Music</b>		<b>Visual Arts</b>	
	<b>Task 1</b>	<b>Task 2</b>	<b>Task 1</b>	<b>Task 2</b>
<i>p</i> -Value	0.71	0.68	0.51	0.49
Discrimination Index	0.57	0.44	0.50	0.37

Johnson, Penny, and Gordon (2008) proposed a discrimination index for performance tasks, which can be used to discriminate between low-performing examinees and high-performing examinees.

It is “an item-criterion correlation, with the criterion being the total raw score on both the multiple-choice and the open-ended items” (p. 271). The Pearson correlation was used to estimate the point-biserial index. The correlations for both music and visual arts performance tasks in 2013 range from .37 to .57. A limited number of studies have been conducted to determine the standard for evaluating discrimination indices for performance tasks. As far as can be determined, only one study by Huynh, Meyer, & Barton (2000) reported an acceptable discrimination index for performance tasks. They suggested that an acceptable index should be around .50. An acceptable discrimination index would indicate that students who do well on a particular performance task tend to score higher on the SCAAP test as a whole than students who perform relatively poorly on that task. Conversely, a low discrimination index means that students who do well on a particular performance task usually score lower on the overall test than students who do not do well in that task. Based on those interpretations, the SCAAP performance tasks for this year function adequately in distinguishing between high-performing and poor-performing students for Tasks 1 in both Music and Visual Arts; however, comparatively, Task 2 for both music and visual arts did not perform as well as Task 1 in both arts areas. As in 2011, the scores for Task 2 in both Music and Visual Arts produced lower values in the discrimination index. In previous year, the discrimination index of Task 2 for both music and visual arts has usually been lower than the discrimination index of Task 1. This pattern suggests further analyses may be helpful in understanding why the indices are decreasing. One of the possibilities may have to do with the task order and student test fatigue. The SCAAP personnel will be working to investigate the possibilities during next year’s assessments.

### ***Internal Validity***

To investigate the internal validity of the SCAAP assessments, correlations across test formats were computed. Specifically, Pearson correlations between students’ multiple-choice test scores and performance task ratings were calculated. 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 5.17: SCAAP 2013 Correlations for Music Multiple-Choice Test Forms and Performance Tasks

	Performance Task 1	Performance Task 2
Multiple-Choice Form 1	0.35	0.20
Multiple-Choice Form 2	0.22	0.25
Performance Task 1	1.00	0.23
Performance Task 2	-	1.00

The Pearson correlations between the various music assessment formats are presented in Table 5.17. 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 correlation between the multiple-choice section and Music Performance Task 1 is .35 for Form 1 and .22 for Form 2. Those correlations indicate the music multiple-choice test and the ratings for Music Performance Task 1 have a shared variance of approximately 5% to 12% ( $r^2$ ). The correlation between each music multiple-choice test forms and Music Performance Task 2 are .20 for Form 1 and .25 for Form 2, indicating that the scores for the music multiple-choice test forms and the scores for Music Performance Task 2 have shared only approximately 4% to 6% variance in common. The moderately low correlations between the multiple-choice test and Music Performance Tasks indicate that the skills required to sing a song and to improvise an 8-beat rhythm pattern are not measured using the SCAAP multiple-choice test format. Because the performance tasks are aligned with specific music standards that are not included in the multiple-choice test forms, this finding validates the inclusion of these two performance tasks in measuring standards-based music achievement.

Table 5.18: *SCAAP 2013 Correlations for Visual Arts Multiple-Choice Test Forms and Performance Tasks*

	Performance Task 1	Performance Task 2
Multiple-Choice Form 1	0.41	0.32
Multiple-Choice Form 2	0.37	0.23
Performance Task 1	1.00	0.18
Performance Task 2	-	1.00

The Pearson correlations between the visual arts multiple-choice and performance task assessment formats are presented in Table 5.18. 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 .41 for Form 1 and .37 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 from 14% to 17% variance in common. The correlation between the Visual Arts multiple-choice test form and Visual Arts Performance Task 2 scores are .32 for Form 1 and .23 for Form 2. This indicates that the Task 2 scores and the multiple-choice scores have a shared variance from about 5% to 10%. The correlations between visual arts multiple-choice test 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 5.18. The two Visual Arts performance tasks share a variance of a little more than 3%, with a correlation of .18. Low correlations were found between Task 1 and Task 2 (.18), suggesting these two tasks are measuring something different in student's visual arts achievement, and it is helpful to keep both tasks in the assessment.

## **VI. RESULTS BY SCHOOL**

The reliability indices for each test were satisfactory for providing school-level results for the SCAAP 2013 music and visual arts assessments. Table 6.1 and 6.2 show the results for the schools that participated in the current year testing. The multiple-choice scores provided are equated scores based on assessment results of the two forms, which were computed using IRT methods. School names are not used in this report to ensure the confidentiality of participating schools. A total of 32 schools participated in the music assessment and 34 schools participated in the visual arts assessment. The number of students per school ranged from 23 to 133 for music and from 23 to 130 for visual arts. The average school score ranged from 13.83 to 31.79 for music and from 16.30 to 35.18 for visual arts. Patterns at the school level are consistent with the overall results. That is, the music assessment tended to be more difficult for students as compared to the visual arts assessment.

Table 6.1: SCAAP 2013 Music Assessment Means by School

School Code	Music Assessment 2012						Music Assessment 2013					
	MC	N	Task 1	N	Task 2	N	MC	N	Task 1	N	Task 2	N
S001	21.71	75	-	-	-	-	22.89	88	8.15	30	5.68	30
S002	22.68	100	-	-	-	-	22.06	104	6.33	15	3.90	13
S003	18.59	46	-	-	-	-	-	-	-	-	-	-
S004	20.59	61	-	-	-	-	-	-	-	-	-	-
S005	20.91	125	-	-	-	-	-	-	-	-	-	-
S006	24.18	70	-	-	-	-	13.83	70	8.35	36	6.38	37
S007	-	-	-	-	-	-	23.11	118	9.21	43	4.09	43
S008	25.28	103	-	-	-	-	-	-	-	-	-	-
S009	30.16	72	-	-	-	-	31.79	70	11.12	40	6.85	40
S010	26.20	93	-	-	-	-	25.95	109	10.94	40	7.66	40
S011	-	-	-	-	-	-	16.75	107	10.27	31	5.97	38
S012	15.47	57	-	-	-	-	13.95	49	-	-	-	-
S013	23.28	106	-	-	-	-	22.89	89	9.48	40	6.02	40
S014	23.22	103	-	-	-	-	-	-	-	-	-	-
S015	-	-	-	-	-	-	28.33	100	9.23	40	4.81	39
S016	22.11	117	-	-	-	-	22.55	133	9.19	40	5.74	40
S017	16.79	57	-	-	-	-	15.48	72	8.72	41	5.21	40
S018	15.46	47	-	-	-	-	16.47	41	7.93	39	5.29	38
S019	17.64	80	-	-	-	-	18.88	67	8.47	40	4.41	39
S020	-	-	-	-	-	-	20.55	72	8.77	40	6.07	40
S021	20.76	110	-	-	-	-	20.74	90	8.53	40	-	-
S022	-	-	-	-	-	-	-	-	-	-	-	-
S023	14.72	102	-	-	-	-	14.84	101	8.58	40	4.89	39
S024	-	-	-	-	-	-	16.85	38	8.41	37	5.08	37
S025	20.04	95	-	-	-	-	27.74	95	10.40	38	7.31	39
S026	13.22	58	-	-	-	-	17.67	70	9.15	39	-	-

School Code	Music Assessment 2012						Music Assessment 2013					
	MC	N	Task 1	N	Task 2	N	MC	N	Task 1	N	Task 2	N
<b>S027</b>	17.82	68	-	-	-	-	17.84	76	10.20	40	-	-
<b>S028</b>	22.79	36	-	-	-	-	24.96	23	10.96	24	7.47	24
<b>S029</b>	26.73	100	-	-	-	-	28.38	107	-	-	-	-
<b>S030</b>	20.55	104	-	-	-	-	21.95	110	10.31	40	6.57	40
<b>S031</b>	23.13	53	-	-	-	-	-	-	-	-	-	-
<b>S032</b>	20.72	49	-	-	-	-	21.45	48	8.78	40	6.03	39
<b>S033</b>	-	-	-	-	-	-	22.14	55	8.08	42	4.90	42
<b>S034</b>	26.35	94	-	-	-	-	-	-	-	-	-	-
<b>S035</b>	-	-	-	-	-	-	28.37	101	9.79	40	6.41	40
<b>S036</b>	-	-	-	-	-	-	22.10	53	8.29	40	6.18	40
<b>S037</b>	-	-	-	-	-	-	25.69	58	9.73	40	5.56	40
<b>S038</b>	29.72	111	-	-	-	-	31.36	129	10.37	40	6.49	39
<b>S039</b>	18.10	42	-	-	-	-	-	-	-	-	-	-
<b>S040</b>	17.20	97	-	-	-	-	17.45	109	11.32	40	6.60	40
<b>S041</b>	20.38	56	-	-	-	-	20.30	47	8.26	42	6.87	42
<b>S042</b>	20.33	58	-	-	-	-	-	-	-	-	-	-



Table 6.2: SCAAP 2013 Visual Arts Assessment Means by School

School Code	Visual Arts Assessment 2012						Visual Arts Assessment 2013					
	MC	N	Task 1	N	Task 2	N	MC	N	Task 1	N	Task 2	N
S001	27.02	74	-	-	-	-	28.98	88	2.39	41	2.98	41
S002	30.48	99	-	-	-	-	29.71	106	2.73	38	2.06	38
S003	21.02	46	-	-	-	-	-	-	-	-	-	-
S004	26.26	61	-	-	-	-	-	-	-	-	-	-
S005	25.40	127	-	-	-	-	-	-	-	-	-	-
S006	24.20	70	-	-	-	-	21.20	70	1.31	41	1.54	41
S007	-	-	-	-	-	-	30.26	115	2.86	40	2.85	40
S008	29.65	111	-	-	-	-	-	-	-	-	-	-
S009	35.71	72	-	-	-	-	35.18	70	3.40	39	2.25	39
S010	25.05	91	-	-	-	-	26.28	113	2.35	40	2.60	40
S011	-	-	-	-	-	-	23.55	114	1.93	40	2.55	40
S012	-	-	-	-	-	-	17.99	49	1.73	40	1.53	40
S013	30.56	106	-	-	-	-	27.97	89	2.21	36	1.42	36
S014	28.69	100	-	-	-	-	-	-	-	-	-	-
S015	-	-	-	-	-	-	33.00	100	1.92	38	2.71	38
S016	30.51	117	-	-	-	-	30.93	130	2.43	36	2.53	36
S017	22.87	61	-	-	-	-	19.66	73	0.97	34	1.81	34
S018	23.68	48	-	-	-	-	25.11	41	2.55	40	1.85	40
S019	25.21	81	-	-	-	-	26.12	70	1.85	41	2.17	41
S020	-	-	-	-	-	-	28.20	72	2.66	40	2.55	40
S021	27.43	114	-	-	-	-	26.17	108	2.52	25	2.68	25
S022	-	-	-	-	-	-	24.67	102	-	-	-	-
S023	18.74	103	-	-	-	-	16.30	100	2.41	39	2.07	39
S024	-	-	-	-	-	-	24.29	39	2.10	38	2.54	38
S025	27.28	95	-	-	-	-	29.65	99	2.98	38	2.07	38
S026	19.36	57	-	-	-	-	20.80	64	1.40	33	1.56	33

School Code	Visual Arts Assessment 2012						Visual Arts Assessment 2013					
	MC	N	Task 1	N	Task 2	N	MC	N	Task 1	N	Task 2	N
<b>S027</b>	23.33	73	-	-	-	-	23.65	77	1.60	40	1.50	40
<b>S028</b>	25.58	37	-	-	-	-	28.19	23	1.84	23	2.13	23
<b>S029</b>	33.49	102	-	-	-	-	32.70	106	-	-	-	-
<b>S030</b>	26.42	103	-	-	-	-	27.78	104	2.43	34	1.74	34
<b>S031</b>	30.30	52	-	-	-	-	-	-	-	-	-	-
<b>S032</b>	28.44	49	-	-	-	-	27.83	47	1.91	37	2.54	37
<b>S033</b>	-	-	-	-	-	-	31.51	54	2.67	40	2.46	40
<b>S034</b>	31.25	94	-	-	-	-	-	-	-	-	-	-
<b>S035</b>	-	-	-	-	-	-	32.57	104	2.46	40	2.18	40
<b>S036</b>	-	-	-	-	-	-	29.45	51	1.68	37	2.41	37
<b>S037</b>	-	-	-	-	-	-	30.18	55	2.22	36	1.78	36
<b>S038</b>	33.99	114	-	-	-	-	34.76	127	2.84	40	1.90	40
<b>S039</b>	22.41	47	-	-	-	-	25.71	27	1.38	25	1.46	25
<b>S040</b>	22.75	99	-	-	-	-	21.76	126	2.82	39	2.55	39
<b>S041</b>	24.14	52	-	-	-	-	22.52	40	1.72	36	1.36	36
<b>S042</b>	25.47	58	-	-	-	-	-	-	-	-	-	-

## VII. TEST ADMINISTRATOR FEEDBACK SURVEY

Test administrators for the 2013 SCAAP music and visual arts assessments were asked to complete an online survey to provide feedback about test administration training procedures and logistics. The following is a synthesis of the results, including test administrators' feedback regarding the training procedures, the multiple-choice assessment, and the performance task assessment.

### Demographics of Survey Respondents

A total of 46 test administrators completed the survey. Of the 46 respondents, 21 (45.7%) administered the music and visual arts assessment respectively. The remaining four test administrators (8.7%) reported that they administered both music and visual arts assessments. Of the 46 respondents, nine administered SCAAP for the first time in 2013, and 10 reported that this was the second time they administered SCAAP. The remaining 27 test administrators had administered SCAAP for at least three years. Further, a majority (n=34) of test administrators who responded to the survey have been teaching for 10 or more years.

### Test Administrator Training Procedures

In 2013, members of the SCAAP team provided a live, interactive test administrator training session through webinar. The webinar training session was also recorded for test administrators to view later at their convenience. Novice SCAAP test administrators were required to attend the training, and experienced SCAAP test administrators were also encouraged to prepare themselves to administer the test using the proper procedures by viewing the training. Of the 46 test administrators who completed the survey, 34.8% (n=16) viewed the training session.

All of the 16 respondents who viewed the training session provided feedback about the quality of the session. When asked to rate the training, two respondents rated the "Ease of accessing the webinar" as "Excellent" and 12 deemed it "Good." One of the remaining two respondents said the training session was "Fair" while the other one rated it "Poor." In terms of the training's "Clarity of content," six test administrators felt it was "Excellent," seven considered it "Good," two respondents rated the clarity of the training as "Fair," and only one replied "Poor." Regarding the "Comprehensiveness" of the training, six respondents felt it was "Excellent," another six test administrators thought it was "Good," three rated it "Fair," and only one respondent deemed it "Poor." With regards to the training's "Usefulness," five test administrators rated it "Excellent,"



another eight rated it “Good.” One of the remaining three respondents assigned the training’s “Usefulness” a rating of “Fair,” while the other two thought it was “Poor.” Table 7.1 displays the percentage of responses on the training webinar for each rating category.

Table 7.1: *Test Administrators’ Rating of Training Session*

	<b>Excellent</b>	<b>Good</b>	<b>Fair</b>	<b>Poor</b>
Ease of accessing webinar	12.5%	75.0%	6.3%	6.3%
Clarity of content	37.5%	43.8%	12.5%	6.3%
Comprehensiveness	37.5%	37.5%	18.8%	6.3%
Usefulness	31.3%	50.0%	6.3%	12.5%

In addition to the specific questions about the training session, test administrators were asked whether they felt that the materials and communications they received from the SCAAP team adequately prepared them to administer the assessment. Of the 16 test administrators who participated the training session, a majority of them (n=14) reported that they were adequately prepared to administer the SCAAP assessments. Only two of them did not think so. One of these two test administrators stated that “Accurate directions for my laptop were not available until I had already completed the testing. I called numerous times at different stages of preparation and still received little help.” A test administrator who did not participate in training stated that some training would be great; he/she experienced some difficulty in administering the SCAAP test because students took the test on laptops.

### **Administering the SCAAP Multiple-Choice Assessments**

One test administrator indicated that the multiple-choice assessment took less than 30 minutes to administer to his/her students. Half of the 46 test administrators reported that it took between 30 minutes and 45 minutes for students to complete this assessment, and 39.1% (n=18) reported that it took 46 to 60 minutes. Four (8.7%) test administrators said it took more than 60 minutes for the students to complete this section.



Overall, 87.0% of the respondents (n=40) indicated that they did not need to provide instructions to students beyond those included in the Test Administrator's Script. Of the six test administrators who did provide additional instructions, one simply reminded students which button to click in order to submit their answer. Another respondent assisted students when a computer froze and the student had to be reassigned to another one. Two test administrators indicated problematic links in the test, and another one pointed out that the bubbles were not aligned with the answers in some questions. This test administrator also said that he/she "reminded students to not listen to all the passages at the same time" and to "scroll down the page to read all possible answers." Finally, a test administrator mentioned that he/she explained how to access the site and how to sign in on the computers; this test administrator also ensured that "each student played the sound clip one at a time."

Approximately 73.9% (n=34) of test administrators reported that their students did not experience technical problems during the multiple-choice assessment. Of those 12 respondents whose students had encountered problems, two indicated that students experienced problems with the sound files and another two incidents were reported for bad links. In other cases, the problems were related to the computers/laptops or the district internet connection (e.g., the computers froze or students kept getting "boot off"). Although test administrators reported problems with sound files, those problems seemed to be less severe than in previous years. One respondent mentioned that some students had to re-enter their username and password, but they had not encountered other issues afterwards. Other technical problems that were reported include the malfunctioning of headphones and the unavailability of the SCAAP test online. The latter problem was solved promptly when the test administrator called the SCAAP office. .

To help the SCAAP team to improve the administration of the multiple-choice assessment, 8.7% of the respondents (n=4) made various suggestions. One test administrator suggested training be provided to him/her while another one suggested training be updated annually. The third test administrator noticed a typo in the administrator's manual, and the last one suggested not allowing students to be able to play the sound files simultaneously because some students were distracted by playing the files together to discover how it would sound.



### **Administering the SCAAP Performance Task Assessments**

A majority of respondents (93.5%, n=43) indicated that they administered the SCAAP performance task assessments. A total of 23 respondents administered the music performance tasks and the remaining 20 administered the visual arts performance tasks. Below is a summary of responses for both the music performance tasks and the visual arts performance tasks.

#### ***The Music Performance Tasks***

Ten test administrators indicated that it took less than five minutes to administer Music Performance Task 1. Another 11 test administrators reported that it took between five and ten minutes for students to complete this task, and only two reported that it took 11 to 20 minutes. For Music Performance Task 2, 10 test administrators reported that it took less than five minutes for students to complete. Twelve respondents indicated this task took between 5 and 10 minutes while only one reported it took between 11 and 20 minutes to administer this task.

#### ***The Visual Arts Performance Tasks***

Only one test administrator indicated that it took between five and ten minutes to administer Visual Arts Performance Task 1. Another three test administrators reported that it took between 11 and 20 minutes for students to complete this task, and 11 reported that it took 21 to 30 minutes. The remaining six respondents said that the task took more than 30 minutes for students to complete. For Visual Arts Performance Task 2, four test administrators reported that it took 11 to 20 minutes to administer the task. Nine respondents indicated that this task took from 21 to 30 minutes. The remaining seven test administrators reported that it took more than 30 minutes to administer this task to students.

Overall, 42 out of the 43 test administrators who administered the performance task assessment provided feedback regarding their experience with this section of the assessment. A total of 36 respondents (85.7%) indicated that they did not need to provide instructions to students beyond those included in the Test Administrator's Script. Of the six test administrators who did provide additional instructions, one simply reminded students to think about their rhythm prior to performance. Another test administrator just simplified the directions because s/he felt the procedure made students nervous. On a similar note, one respondent indicated students were unclear about how the paper could be oriented in the Visual Arts Performance

Task 2: Drawing. Furthermore, a test administrator mentioned that s/he provided draft paper for students for writing. The remaining replies said “the prompts did not provide clear downbeats” and that the word bank might not be necessary because his/her students copied words from the work bank instead of providing an explanation.

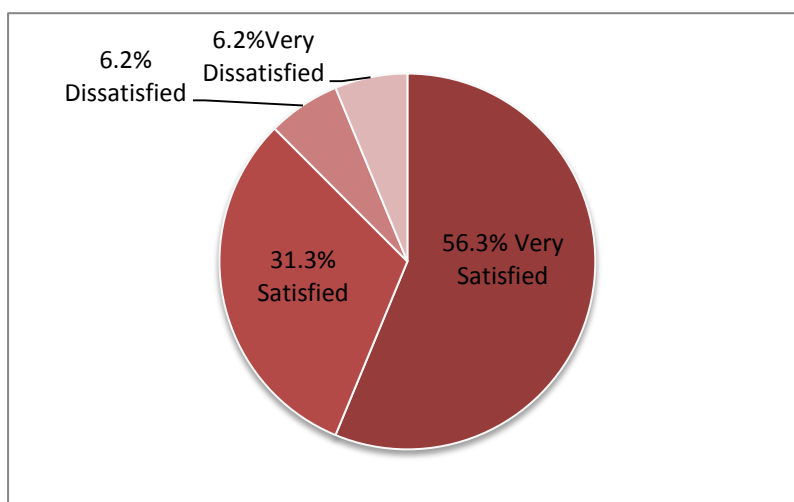
The majority of test administrators (76.2%, n=32) did not provide any suggestions on how to improve the administration procedures for the performance tasks. Of those 10 respondents who did so, four focused on the test instructions. One mentioned the instructions were lengthy, and another one said he/she had to replay the rhythm instructions to help students understand. In terms of writing instructions, a respondent suggested that they be more like PASS test. The last suggestion for the task instructions is to make them available for all types of computer systems. Besides test instructions, suggestions centered around examples or test prompts. One test administrator proposed to “leave out the rhythmic example or change the rhythm between the example and the test” because students tended to get the “rhythmic idea stuck in their head.” Another test administrator felt that rhythmic prompts should be in the same meter as expected in the rhythmic improvisation. A third test administrator suggested clearer examples for the compare and contrast task. One test administrator noted that making two separate files for singing and rhythm takes more time but acknowledged that they are scored separately.

When asked how test administrators accessed the music prompts when administering the music performance tasks, only three test administrators mentioned using the digital files provided on the flash drive. The remaining test administrators all used the CD provided to them: five played it through their computers while 16 played it using a separate CD player. Among these test administrators, only one of them reported that s/he had experienced difficulty in timing the recording and audio appropriately. This test administrator planned to use a separate audio device in future testing. One test administrator put the files on iTunes and reported that the procedure worked well. Another noted that test administration was time-consuming even though s/he did not encounter any difficulties.

Despite various technical problems encountered, less than half of the 46 test administrators (34.8%, n=16) interacted with the SCAAP personnel either through email or phone call during



the actual administration of the SCAAP test. In addition, a majority of the respondents reported a strong degree of satisfaction with the SCAAP team's timely support and prompt response. Of those who responded to this part of the survey, 87.5% reported they were either "Very Satisfied" or "Satisfied" with the support and assistance they had received from the SCAAP team. Only one out of the 16 test administrators said he/she was not satisfied and another one expressed strong dissatisfaction. This test administrator experienced difficulty in SCAAP testing due to the use of a laptop that ran on a different system. After the information was communicated to the SCAAP team, the team designed a test manual especially for this particular school. This manual was immediately sent to the test administrator. Below, Figure 7.1 displays respondents' satisfaction with the SCAAP team; these results are consistent with those from the 2012-2013 testing cycle.



*Figure 7.1. Response to: Rate your level of satisfaction with the support and assistance you received from the SCAAP team during test administration.*

Among the 10 test administrators who provided additional comments, six were related to the SCAAP testing. One respondent reacted positively to the overall testing procedure, while another replied that he/she did not see the benefit of the performance tasks. While two test administrators asked about the possibility of receiving individual feedback/analysis on student performance, one mentioned that his/her students needed something to help them study for the test because some of the content was not previously covered. Finally, one respondent supplied a detailed list of suggestions regarding the appropriateness of testing in the arts.



## VIII. CONCLUSIONS

The section summarizes findings from the 2013 administration of the SCAAP assessments. Highlights of key activities contributing to SCAAP assessment development that occurred over the 2012-2013 academic year are also discussed.

### Technical Characteristics

Using Cronbach's Alpha, the reliability ranged from .79-.82 on the two forms of the music assessment and from .85-.86 on the two forms of the visual arts assessment. These are acceptable levels of reliability for medium-stakes decisions. The reliability at the standard level had a lower range; .16 to .66 for music standards with two to 18 items included and .25 to .65 for visual arts standards with four to 11 items included. The standards with relatively fewer items had the lowest reliability, which is to be expected.

The  $p$ -value, or proportion correct, measures the level of difficulty of an item. The  $p$ -values ranged from .27 to .88 on the music assessment and from .06 to .81 on the visual arts assessment. One visual arts item was considered extremely difficult with  $p$ -values below .2. Seven music items and two visual arts items were considered extremely easy with  $p$ -values above .8.

Ten music items and two visual arts items were identified for relatively low discrimination between high and low performing students and will be reviewed in the 2013-2014 year. In addition, two music items were found to exhibit DIF between white and African American students and will also be reviewed during the 2013-2014 project year. Based on feedback from Arts Advisors, the items identified for review may be revised or discontinued for use on future assessments.

### Student Performance

A total of 2,599 students from 32 schools completed to the SCAAP music assessment and 2,763 students from 34 schools completed the SCAAP visual arts assessment in 2013. Results indicate that students performed better, on average, on the visual arts assessment than the music assessment, though there was greater variation among visual arts scores. This pattern is consistent with previous administrations of the SCAAP assessments.



### **Highlights of 2012-2013 SCAAP Development**

Due to budget constraints, the 2012 SCAAP assessments only included the multiple-choice section. In 2013, the performance assessment of both music and visual arts for the fourth grade resumed to the administration. Even though there was a one-year gap between the administrations of the performance tasks, the administration ran smoothly. The SCAAP personnel at the OPE developed two different versions of the music test administrator manuals to meet the needs of teachers using different computer systems.

In previous years, the music prompts and instructions were usually burned into CDs and provided to music teachers for music performance tasks. This year, the music prompts and instruction files were sent by CDs and also provided on the flash drive that the test administrators use to record students' performances. Sending the digital files on the flash drive allows music test administrators to use multiple electronic devices to play the prompts and instructions for students.



## 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.
- DeVellis, R.F. (1991). *Scale Development*. Newbury Park, NJ: Sage Publications.
- 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.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park: Sage.
- 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. (1988). Equating minimum-competency tests: Comparisons of methods. *Journal of Educational Measurement* 25, 221-231.
- Hill, R. (2002, April). Examining the reliability of accountability systems. Presented at the Annual Meeting of the American Educational Research Association, New Orleans.
- Huynh, H., Meyer, P., & Barton, K. (2000). *Technical documentation for the 1999 Palmetto Achievement Challenge Tests of English language arts and mathematics, grades three through eight*. Columbia: University of South Carolina.
- Johnson, R.L., Penny, J., Fisher, S., & Kuhs, T. (2000). The relationship between score resolution methods and inter-rater 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 inter-rater 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.



- Kaplan, R. & Saccuzzo, D. (1982). *Psychological Testing: Principles, applications, and issues*. Monterey, CA: Brooks/Cole.
- Kuhs, T.M., Johnson, R.L., Agruso, S.A., & Monrad, D.M. (2001). *Put to the Test: Tools and Techniques for Classroom Assessment*. Portsmouth, NH: Heinemann.
- 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.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology (4<sup>th</sup> ed)*. Orlando, FL: Harcourt Brace Jovanovich.
- Nunally, J. (1978). *Psychometric theory (2<sup>nd</sup> ed.)*. New York: McGraw-Hill.
- Penny, J., Johnson, R.L., Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education*, 68(3), 269-288.
- Petersen, N.S., Cook, L.L., & Stocking, M.L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8, 137-156.
- Phillips, S. E. (2000). "Legal Corner: GI Forum v TEA." *NCME Newsletter* 8, no.2 (April): n.p. SAS Institute (1990). *SAS User's Guide*. Cary, NC: Author
- SCDE (1999). *South Carolina Visual and Performing Arts Standards*. SC: South Carolina Department of Education.
- SCDE (2002). *South Carolina Visual and Performing Arts Curriculum Standards 2002 Field Review Draft*. SC: South Carolina Department of Education.
- SCDE (2003). *South Carolina Visual and Performing Arts Curriculum Standards 2003*. SC: South Carolina Department of Education.
- SCDE (2010). *South Carolina Academic Standards for the Visual and Performing Arts*. SC: South Carolina Department of Education.
- Shavelson, R.J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- Yap, C.C., Schneider, C., Johnson, R.L., Mazzie, D., & Porchea, S.(2002). *Technical Documentation for the South Carolina Arts Assessment Project (SCAAP) Year 1: 4<sup>th</sup> Grade Music and Visual Arts Assessments*. Prepared for the South Carolina State

Department of Education by the Office of Program Evaluation, University of South Carolina.

- Yap, C.C., Johnson, R.L., Brown, L.B., Moore, M., & Yoshioka, Y. (2003). *Technical Documentation for the South Carolina Arts Assessment Program (SCAAP) Year 2: 4<sup>th</sup> Grade Music and Visual Arts Assessments*. Prepared for the South Carolina State Department of Education by the Office of Program Evaluation, University of South Carolina.
- Yap, C.C., Moore, M., & Peng, P. (2004). *Technical Documentation for the South Carolina Arts Assessment Program (SCAAP) Year 3: 4<sup>th</sup> Grade Music and Visual Arts Assessments*. Prepared for the South Carolina State Department of Education by the Office of Program Evaluation, University of South Carolina.
- 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.

### Appendix A: 2012 SCAAP BILOG Syntax

```

>GLOBAL DFName = 'F:\Equating_VA2012\VA2012joint.txt',
    NPArm = 2,
    LOGistic,
    OMIIts,
    SAVe;
>SAVE MAsTer = 'VA2012.MAS',
    PARm = 'VA2012.PAR',
    SCORe = 'VA2012.SCO',
    EXPeCted = 'VA2012.EXP';
>LENGTH NITems = (64);
>INPUT NTOtal = 64,
    NALt = 4,
    NIDchar = 5,
    KFName = 'F:/Equating_VA2012/VA2012jointkey.txt',
    NFName = 'F:/Equating_VA2012/VA2012jointkey2.txt',
    OFName = 'F:/Equating_VA2012/VA2012jointkey3.txt';
>ITEMS INAmes = (ITEM101, ITEM102, ITEM103, ITEM104, ITEM105, ITEM106, ITEM107,
ITEM108, ITEM109, ITEM110, ITEM111, ITEM112, ITEM113, ITEM114, ITEM115, ITEM116,
ITEM117, ITEM118, ITEM119, ITEM120, ITEM121, ITEM122, ITEM123, ITEM124, ITEM125,
ITEM126, ITEM127, ITEM128, ITEM129, ITEM130, ITEM131, ITEM132, ITEM133, ITEM134,
ITEM135, ITEM136, ITEM137, ITEM138, ITEM139, ITEM140, ITEM141, ITEM142, ITEM143,
ITEM144, ITEM145, ITEM204, ITEM208, ITEM211, ITEM213, ITEM214, ITEM219, ITEM220,
ITEM226, ITEM228, ITEM229, ITEM232, ITEM233, ITEM234, ITEM235, ITEM236, ITEM237,
ITEM238, ITEM240, ITEM245 );
>TEST1 TNAmE = 'TEST0001',
    INUmber = (1(1)64);
(5A1, 13X, 64A1)
>CALIB CYCles = 40,
    NEWton = 100,
    ACCel = 1.0000;
>SCORE ;

```



### Appendix B: Supplemental Tables

Table B1: SCAAP 2013 Descriptive Statistics of Test Scores by Ethnicity

Content Area	Form	Ethnicity											
		Asian			Native American			Other			Two or More Race Categories		
		M	SD	n	M	SD	n	M	SD	n	M	SD	n
Music	1	26.9	7.1	12	19.5	2.1	2	22.5	12.4	4	23.4	6.1	25
	2	24.6	7.3	6	30.0	-	1	20	5.6	2	23.1	6.4	39
Visual Arts	1	29.5	6.7	15	20.5	0.7	2	-	-	-	26.5	8.0	31
	2	29.1	6.7	6	27.5	0.7	2	22.7	6.9	7	27.0	7.2	39

Table B2: SCAAP 2013 Ranges of P-values of Test Forms by Ethnic Groups

Content Area	Form	P-value Ranges			
		Asian	Native American	Other	Two or More Race Categories
Music	1	0.25-1.00	0.00-1.00	0.00-1.00	0.12-1.00
	2	0.00-1.00	0.00-1.00	0.00-1.00	0.15-0.90
Visual Arts	1	0.07-1.00	0.00-1.00	-	0.10-0.90
	2	0.00-1.00	0.00-1.00	0.00-1.00	0.15-0.92

Table B3 : 2013 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	0.56	0.38	22	18	2	56*	A	A
2	0.62	0.46	26	3	61*	8	B	A
3	0.83	0.36	82*	6	5	4	A	A
4	0.41	0.18	40	9	9	41*	A	A
5	0.60	0.44	26	3	59*	9	B	A
6	0.47	0.37	17	11	24	47*	A	B
7	0.30	0.19	26	30*	23	19	A	A
8	0.33	0.18	19	21	32*	26	A	B
9	0.51	0.32	7	15	25	51*	A	A
10	0.85	0.38	8	3	3	84*	A	A
11	0.31	0.32	31*	29	21	17	A	A
12	0.45	0.21	12	45*	11	30	A	A
13	0.48	0.39	11	48*	12	28	A	A
14	0.42	0.18	29	19	42*	8	A	A
15	0.31	0.34	7	29	30*	31	A	A
16	0.46	0.36	10	45*	27	16	A	A
17	0.33	0.22	16	30	20	33*	A	A
18	0.39	0.35	20	39*	18	21	A	A
19	0.45	0.29	45*	20	20	14	A	A
20	0.41	0.44	41*	27	14	16	A	A
21	0.64	0.18	13	63*	12	10	A	B
22	0.36	0.26	50	36*	7	5	A	A
23	0.35	0.10	35*	29	25	9	A	A
24	0.36	0.21	12	34	36*	16	A	A
25	0.43	0.38	23	15	42*	18	A	A
26	0.42	0.37	25	18	42*	13	A	A
27	0.39	0.23	22	23	38*	15	A	A
28	0.35	0.35	7	34*	32	24	A	B



Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			A	B	C	D		
29	0.53	0.25	20	53*	19	6	A	A
30	0.50	0.46	14	21	12	50*	A	A
31	0.41	0.35	22	41*	14	21	A	A
32	0.72	0.46	10	8	8	72*	A	A
33	0.28	0.22	15	47	27*	9	A	A
34	0.47	0.39	16	47*	11	24	A	A
35	0.51	0.20	51*	15	9	23	A	A
36	0.68	0.39	9	5	68*	16	A	A
37	0.47	0.31	16	12	47*	23	B	A
38	0.38	0.33	38*	25	11	24	A	A
39	0.29	0.26	29	29	10	29*	A	A
40	0.86	0.40	4	3	4	87*	B	B
41	0.77	0.34	2	77*	5	14	A	A
42	0.86	0.39	3	86*	4	4	A	B
43	0.39	0.36	8	10	42	39*	A	A
44	0.33	0.31	33*	9	30	26	A	A
45	0.33	0.19	28	22	34*	15	A	A

Table B4: 2013 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	0.77	0.36	13	76*	3	6	A	A
2	0.53	0.33	9	52*	33	5	A	A
3	0.76	0.39	9	4	75*	10	A	A
4	0.50	0.49	32	4	49*	13	A	A
5	0.63	0.48	18	62*	9	9	A	A
6	0.33	0.30	11	33*	31	23	A	A
7	0.46	0.20	40	45*	7	6	A	C
8	0.34	0.18	19	22	34*	23	A	A
9	0.57	0.45	14	9	57*	18	A	A
10	0.28	0.17	30	28*	21	19	A	A
11	0.86	0.38	8	3	2	86*	A	C
12	0.40	0.23	12	40*	11	36	A	A
13	0.36	0.37	36*	24	21	17	A	A
14	0.52	0.35	9	52*	12	25	A	A
15	0.45	0.18	30	17	44*	7	A	B
16	0.47	0.40	11	20	47*	20	A	A
17	0.37	0.34	14	26	21	37*	A	A
18	0.42	0.37	16	42*	19	21	A	A
19	0.47	0.29	47*	18	19	13	A	A
20	0.42	0.40	42*	28	12	16	A	B
21	0.60	0.21	13	59*	15	11	A	B
22	0.78	0.26	77*	4	10	7	A	B
23	0.41	0.30	44	41*	7	7	A	A
24	0.35	0.15	35*	29	26	9	A	A
25	0.40	0.36	23	20	39*	15	A	A
26	0.40	0.36	40*	24	17	17	A	A
27	0.63	0.41	12	63*	13	10	A	A
28	0.39	0.37	8	39*	32	19	A	A
29	0.34	0.37	35	33*	11	18	A	A
30	0.41	0.39	30	41*	17	10	A	A
31	0.63	0.48	63*	14	6	15	B	A

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			A	B	C	D		
32	0.61	0.38	19	61*	10	7	A	A
33	0.44	0.27	44*	24	6	23	A	A
34	0.66	0.41	66*	10	12	11	A	A
35	0.37	0.37	30	23	8	37*	A	A
36	0.59	0.38	24	5	10	59*	A	A
37	0.43	0.46	23	43*	17	16	A	B
38	0.48	0.35	47*	21	8	22	A	A
39	0.27	0.26	30	29	12	26*	A	A
40	0.88	0.37	2	4	3	89*	B	B
41	0.77	0.34	1	77*	6	15	A	A
42	0.87	0.34	2	87*	4	5	A	B
43	0.65	0.49	17	9	8	65*	A	A
44	0.34	0.27	34*	11	27	27	A	A
45	0.35	0.18	28	20	35*	16	A	A

Table B5: 2013 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	0.77	0.39	1	15	5	77*	A	A
2	0.73	0.29	8	73*	8	10	A	A
3	0.22	0.20	21	21*	14	42	A	A
4	0.73	0.39	72*	10	8	8	A	A
5	0.76	0.42	2	15	6	75*	A	A
6	0.42	0.39	41*	15	28	13	A	A
7	0.66	0.40	8	13	11	65*	A	A
8	0.69	0.26	11	7	11	69*	A	B
9	0.74	0.42	11	73*	7	7	A	A
10	0.49	0.31	23	17	49*	10	A	A
11	0.65	0.33	65*	10	18	5	A	A
12	0.69	0.47	10	69*	7	12	A	A
13	0.37	0.27	14	27	37*	21	A	A
14	0.40	0.38	40*	14	20	24	A	A
15	0.72	0.46	13	72*	7	7	A	A
16	0.48	0.33	29	47*	10	12	A	A
17	0.66	0.25	15	3	65*	15	A	A
18	0.62	0.50	14	7	15	62*	A	A
19	0.61	0.41	16	11	11	60*	A	A
20	0.78	0.32	8	5	77*	8	A	A
21	0.52	0.38	11	8	51*	28	A	A
22	0.53	0.30	13	21	11	53*	A	A
23	0.33	0.48	32*	33	17	16	A	A
24	0.25	0.32	24*	29	19	25	A	A
25	0.80	0.32	80*	10	4	5	A	A
26	0.12	0.06	12*	17	13	55	A	A
27	0.26	0.29	25	24	26*	23	A	A
28	0.67	0.38	16	67*	5	10	A	A
29	0.73	0.13	3	17	72*	6	A	B
30	0.41	0.36	18	15	24	41*	A	A
31	0.55	0.45	23	55*	14	6	A	A

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			16	29*	28	24		
32	0.30	0.27	16	29*	28	24	A	A
33	0.80	0.40	4	80*	8	6	B	A
34	0.51	0.36	19	8	51*	19	A	A
35	0.56	0.50	7	21	55*	14	A	A
36	0.62	0.33	28	4	4	62*	A	A
37	0.45	0.44	12	31	45*	9	A	A
38	0.60	0.53	15	12	11	60*	A	A
39	0.65	0.33	7	64*	19	8	A	A
40	0.39	0.40	39*	28	13	18	A	A
41	0.68	0.44	4	18	68*	8	A	B
42	0.52	0.51	52*	17	14	15	A	A
43	0.65	0.50	65*	9	19	5	A	A
44	0.68	0.49	68*	11	14	5	A	A
45	0.41	0.36	23	41*	21	14	A	A

Table B6: 2013 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	0.81	0.36	2	10	5	80*	A	A
2	0.73	0.26	8	73*	9	9	B	A
3	0.39	0.41	39*	16	31	13	A	A
4	0.72	0.40	71*	9	8	9	A	A
5	0.74	0.41	3	15	7	73*	A	A
6	0.21	0.22	21	20*	13	44	A	A
7	0.81	0.44	4	5	8	80*	A	A
8	0.68	0.26	10	9	11	68*	A	B
9	0.53	0.27	19	18	8	53*	A	A
10	0.45	0.33	26	18	45*	9	A	A
11	0.59	0.32	14	19	7	58*	A	A
12	0.69	0.43	11	69*	8	10	A	A
13	0.38	0.44	20	38*	20	20	A	A
14	0.42	0.35	42*	13	22	22	A	A
15	0.33	0.27	15	32	18	32*	A	A
16	0.48	0.39	27	48*	10	14	A	A
17	0.53	0.37	20	11	14	53*	A	A
18	0.61	0.39	61*	14	18	5	A	A
19	0.64	0.44	14	9	12	64*	A	A
20	0.73	0.41	16	4	5	73*	A	A
21	0.74	0.51	11	74*	6	7	A	A
22	0.28	0.26	34	11	27*	26	A	A
23	0.39	0.35	38*	11	29	20	A	A
24	0.49	0.43	26	48*	18	6	A	A
25	0.62	0.46	11	62*	14	12	A	A
26	0.33	0.44	32*	32	17	16	A	A
27	0.64	0.51	8	64*	14	13	A	A
28	0.53	0.44	23	9	14	53*	A	A
29	0.52	0.39	9	8	52*	29	A	A
30	0.72	0.39	72*	12	10	4	A	A
31	0.54	0.43	24	54*	13	7	A	A

Item	P-value	Discrimination Index	Percentage of Options				Gender DIF	Ethnic DIF
			18	30*	27	23		
32	0.30	0.25	18	30*	27	23	A	B
33	0.55	0.49	14	15	55*	14	A	A
34	0.44	0.22	21	7	44*	27	A	A
35	0.42	0.30	19	18	19	42*	A	A
36	0.60	0.43	13	14	60*	11	A	A
37	0.44	0.24	34	11	9	44*	A	A
38	0.53	0.36	13	53*	22	10	A	A
39	0.49	0.30	8	49*	25	17	A	A
40	0.77	0.38	6	77*	9	5	A	A
41	0.66	0.45	4	17	65*	12	A	A
42	0.47	0.46	46*	19	13	20	A	A
43	0.60	0.34	28	5	6	60*	A	A
44	0.25	0.30	24	21	25*	28	A	A
45	0.64	0.45	64*	9	19	5	A	A

Table B7: 2013 SCAAP Item P-Values for Music Test Forms by Ethnicity

Item	Music Form 1			Music Form 2		
	African-American	White	ETS $\Delta$	African-American	White	ETS $\Delta$
1	0.49	0.63	0.03	0.71	0.81	0.17
2	0.53	0.71	0.03	0.51	0.55	0.79
3	0.77	0.90	-0.64	0.72	0.80	0.68
4	0.41	0.41	0.84	0.44	0.57	0.57
5	0.54	0.67	0.47	0.52	0.74	-0.91
6	0.36	0.58	-1.03	0.30	0.38	0.20
7	0.26	0.33	-0.21	0.51	0.41	1.81
8	0.34	0.32	1.04	0.34	0.35	0.47
9	0.49	0.55	0.72	0.47	0.69	-0.75
10	0.78	0.91	-0.72	0.26	0.31	-0.07
11	0.25	0.38	-0.44	0.78	0.94	-2.13
12	0.46	0.45	0.84	0.40	0.42	0.55
13	0.41	0.55	0.22	0.30	0.42	-0.08
14	0.42	0.45	0.37	0.48	0.56	0.47
15	0.23	0.38	-0.48	0.48	0.41	1.37
16	0.38	0.54	-0.24	0.39	0.56	-0.44
17	0.34	0.34	0.98	0.34	0.43	0.30
18	0.31	0.48	-0.54	0.36	0.49	-0.06
19	0.43	0.49	0.58	0.44	0.51	0.36
20	0.32	0.52	-0.45	0.33	0.55	-1.01
21	0.67	0.62	1.62	0.62	0.58	1.28
22	0.34	0.41	0.20	0.77	0.77	1.26
23	0.36	0.34	0.75	0.37	0.45	0.20
24	0.35	0.38	0.36	0.36	0.35	0.65
25	0.35	0.52	-0.34	0.33	0.48	-0.20
26	0.37	0.49	0.18	0.36	0.45	0.27
27	0.39	0.40	0.89	0.55	0.72	-0.46
28	0.25	0.45	-1.16	0.32	0.49	-0.58
29	0.44	0.61	-0.66	0.28	0.40	0.03
30	0.42	0.59	0.01	0.34	0.48	-0.12
31	0.32	0.51	-0.83	0.54	0.72	-0.42
32	0.62	0.82	-0.60	0.56	0.68	0.18
33	0.22	0.34	-0.55	0.38	0.52	-0.62
34	0.38	0.56	-0.55	0.59	0.75	-0.14
35	0.48	0.55	0.17	0.29	0.47	-0.70
36	0.60	0.74	-0.11	0.50	0.68	-0.62



Item	Music Form 1			Music Form 2		
	African-American	White	ETS $\Delta$	African-American	White	ETS $\Delta$
37	0.38	0.57	-0.72	0.30	0.57	-1.47
38	0.32	0.44	0.14	0.40	0.56	-0.51
39	0.25	0.34	0.00	0.21	0.32	-0.37
40	0.78	0.94	-1.43	0.83	0.94	-1.18
41	0.71	0.83	-0.32	0.70	0.83	-0.63
42	0.78	0.93	-1.15	0.80	0.93	-1.49
43	0.33	0.45	0.03	0.53	0.75	-0.82
44	0.28	0.39	-0.16	0.31	0.36	0.50
45	0.32	0.36	0.21	0.32	0.36	0.18

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

Table B8: 2013 SCAAP Item P-Values for Visual Arts Test Forms by Ethnicity

Item	Visual Arts Form 1			Visual Arts Form 2		
	African-American	White	ETS $\Delta$	African-American	White	ETS $\Delta$
1	0.72	0.82	0.55	0.75	0.85	0.55
2	0.67	0.80	-0.39	0.67	0.80	-0.52
3	0.17	0.29	-0.93	0.31	0.49	0.03
4	0.66	0.79	0.17	0.65	0.78	0.36
5	0.68	0.83	0.00	0.65	0.84	-0.75
6	0.35	0.51	0.13	0.18	0.24	0.41
7	0.59	0.74	0.38	0.74	0.88	0.08
8	0.67	0.69	1.18	0.67	0.70	1.04
9	0.66	0.82	-0.27	0.49	0.59	0.30
10	0.40	0.60	-0.69	0.38	0.54	-0.14
11	0.59	0.72	0.39	0.49	0.68	-0.67
12	0.59	0.79	-0.20	0.61	0.76	0.47
13	0.32	0.45	-0.21	0.29	0.49	-0.16
14	0.33	0.49	0.40	0.36	0.48	0.53
15	0.65	0.81	0.39	0.31	0.37	0.87
16	0.41	0.55	0.14	0.42	0.54	0.48
17	0.63	0.68	0.74	0.46	0.62	-0.05
18	0.51	0.74	-0.40	0.52	0.71	-0.36
19	0.54	0.69	0.33	0.58	0.70	0.79
20	0.70	0.85	-0.86	0.66	0.80	0.18
21	0.40	0.64	-0.78	0.65	0.85	-0.29
22	0.47	0.59	0.18	0.22	0.35	-0.47
23	0.24	0.42	0.59	0.33	0.46	0.26
24	0.18	0.32	-0.07	0.37	0.61	-0.66
25	0.75	0.86	-0.11	0.52	0.75	-0.61
26	0.15	0.11	0.29	0.22	0.45	-0.59
27	0.18	0.34	-0.63	0.52	0.77	-0.72
28	0.58	0.76	-0.26	0.42	0.66	-0.70
29	0.74	0.71	1.16	0.44	0.59	0.34
30	0.34	0.49	0.21	0.64	0.82	-0.68
31	0.44	0.68	-0.46	0.41	0.66	-0.85
32	0.26	0.33	0.56	0.29	0.32	1.08
33	0.75	0.85	0.30	0.43	0.70	-0.65
34	0.43	0.61	-0.44	0.41	0.47	0.32
35	0.43	0.69	-0.47	0.34	0.51	-0.51
36	0.57	0.68	0.32	0.51	0.72	-0.55

Item	Visual Arts Form 1			Visual Arts Form 2		
	African-American	White	ETS $\Delta$	African-American	White	ETS $\Delta$
37	0.36	0.57	-0.07	0.37	0.51	-0.61
38	0.46	0.75	-0.90	0.47	0.61	0.13
39	0.55	0.75	-0.69	0.43	0.54	0.26
40	0.28	0.51	-0.57	0.72	0.82	0.84
41	0.61	0.74	1.02	0.6	0.72	0.93
42	0.43	0.62	0.61	0.37	0.58	-0.20
43	0.53	0.77	-0.33	0.56	0.65	0.57
44	0.59	0.78	0.25	0.22	0.31	0.62
45	0.32	0.51	-0.40	0.55	0.74	-0.13

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

Table B9: 2013 SCAAP Item P-Values for Music Test Forms by Gender

Item	Music Form 1			Music Form 2		
	Female	Male	ETS $\Delta$	Female	Male	ETS $\Delta$
1	0.55	0.58	0.69	0.79	0.74	0.37
2	0.58	0.65	1.34	0.51	0.54	0.25
3	0.86	0.80	-0.91	0.75	0.76	-0.72
4	0.42	0.40	-0.10	0.50	0.49	0.03
5	0.55	0.64	1.50	0.64	0.62	-0.23
6	0.49	0.45	-0.05	0.37	0.29	-0.55
7	0.33	0.28	-0.49	0.46	0.45	0.25
8	0.33	0.32	0.02	0.37	0.32	0.68
9	0.56	0.47	-0.63	0.62	0.53	-0.31
10	0.87	0.82	-0.85	0.28	0.29	-0.22
11	0.30	0.32	0.54	0.86	0.86	-0.06
12	0.48	0.43	-0.31	0.43	0.37	0.26
13	0.50	0.46	-0.19	0.39	0.33	0.50
14	0.45	0.39	-0.61	0.55	0.50	0.03
15	0.31	0.31	0.23	0.44	0.45	0.05
16	0.46	0.45	0.19	0.47	0.47	-0.50
17	0.32	0.34	0.23	0.39	0.36	0.59
18	0.42	0.36	-0.30	0.44	0.40	-0.16
19	0.47	0.42	-0.34	0.51	0.44	-0.89
20	0.39	0.44	0.93	0.42	0.43	0.95
21	0.64	0.63	-0.07	0.62	0.58	-0.40
22	0.34	0.38	0.64	0.81	0.74	0.28
23	0.37	0.33	-0.49	0.38	0.44	-0.65
24	0.36	0.36	0.10	0.38	0.33	-0.67
25	0.46	0.39	-0.56	0.40	0.40	0.31
26	0.40	0.44	0.77	0.45	0.36	0.67
27	0.39	0.38	0.11	0.68	0.59	0.19
28	0.35	0.35	0.23	0.40	0.39	-1.35
29	0.52	0.54	0.24	0.33	0.35	0.27
30	0.51	0.50	0.15	0.42	0.40	0.90
31	0.40	0.42	0.43	0.71	0.56	0.92
32	0.73	0.71	0.02	0.62	0.61	0.22
33	0.24	0.31	0.94	0.41	0.47	0.44
34	0.44	0.50	0.90	0.64	0.68	0.47
35	0.52	0.50	-0.05	0.38	0.37	0.37
36	0.71	0.64	0.69	0.59	0.59	0.25
37	0.52	0.42	1.34	0.43	0.43	-0.72

Item	Music Form 1			Music Form 2		
	Female	Male	ETS $\Delta$	Female	Male	ETS $\Delta$
38	0.41	0.34	-0.59	0.50	0.45	-0.35
39	0.30	0.28	-1.06	0.29	0.24	-0.46
40	0.90	0.82	-0.57	0.91	0.86	-1.00
41	0.80	0.73	-0.15	0.79	0.75	-0.42
42	0.87	0.85	-1.52	0.89	0.84	-0.56
43	0.39	0.38	-0.88	0.67	0.62	-0.02
44	0.34	0.31	-0.25	0.37	0.31	-0.48
45	0.35	0.32	-0.08	0.37	0.33	-0.28

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

Table B10: 2013 SCAAP Item P-Values for Visual Arts Test Forms by Gender

Item	Visual Arts Form 1			Visual Arts Form 2		
	Female	Male	ETS $\Delta$	Female	Male	ETS $\Delta$
1	0.79	0.76	0.38	0.80	0.81	-0.09
2	0.77	0.68	-0.71	0.78	0.68	0.06
3	0.24	0.19	-0.47	0.41	0.37	-0.43
4	0.74	0.71	0.30	0.74	0.69	-0.34
5	0.79	0.72	-0.21	0.76	0.72	-0.53
6	0.43	0.40	0.32	0.23	0.19	-0.35
7	0.71	0.60	-0.58	0.84	0.78	0.16
8	0.74	0.64	-0.89	0.72	0.65	0.15
9	0.77	0.70	-0.27	0.56	0.50	0.70
10	0.51	0.48	0.16	0.46	0.44	0.76
11	0.64	0.67	0.92	0.60	0.58	-0.22
12	0.70	0.67	0.50	0.69	0.69	0.00
13	0.40	0.34	-0.38	0.36	0.39	0.01
14	0.44	0.35	-0.46	0.44	0.40	-0.28
15	0.76	0.68	-0.20	0.34	0.32	1.00
16	0.51	0.43	-0.38	0.50	0.46	-0.95
17	0.68	0.63	-0.23	0.56	0.50	0.42
18	0.68	0.56	-0.66	0.59	0.64	-0.38
19	0.65	0.55	-0.52	0.70	0.59	-0.34
20	0.79	0.77	0.26	0.74	0.73	-0.40
21	0.55	0.47	-0.32	0.78	0.71	0.40
22	0.55	0.51	0.04	0.30	0.26	0.10
23	0.35	0.30	0.20	0.42	0.36	0.45
24	0.24	0.26	0.97	0.49	0.48	-0.26
25	0.81	0.79	0.26	0.64	0.60	0.63
26	0.12	0.13	0.08	0.33	0.33	0.13
27	0.29	0.22	-0.54	0.67	0.61	0.60
28	0.71	0.62	-0.55	0.53	0.54	0.45
29	0.71	0.74	0.64	0.53	0.51	-0.11
30	0.42	0.39	0.17	0.72	0.72	-0.59
31	0.57	0.54	0.36	0.54	0.54	-0.49
32	0.29	0.30	0.59	0.32	0.29	0.61
33	0.84	0.74	-1.01	0.60	0.51	-0.61
34	0.54	0.48	-0.01	0.47	0.40	-0.31
35	0.59	0.52	-0.08	0.40	0.44	-0.09
36	0.64	0.61	0.26	0.65	0.56	0.06
37	0.47	0.43	0.22	0.47	0.41	-0.43

Item	Visual Arts Form 1			Visual Arts Form 2		
	Female	Male	ETS $\Delta$	Female	Male	ETS $\Delta$
38	0.64	0.55	-0.12	0.54	0.52	0.22
39	0.65	0.64	0.40	0.51	0.47	-0.15
40	0.41	0.36	0.13	0.78	0.76	0.16
41	0.70	0.67	0.41	0.70	0.62	-0.46
42	0.54	0.50	0.50	0.46	0.47	0.53
43	0.69	0.60	-0.32	0.61	0.59	0.11
44	0.71	0.65	0.17	0.24	0.27	0.76
45	0.41	0.40	0.55	0.68	0.61	-0.46

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