Developing a Bundled Visual Arts Assessment Model

Thomas M. Brewer, Ph.D.
University of Central Florida

As art educators, we should be past the point of being afraid of or threatened by assessment. As a discipline we need it, but two persistent questions remain: What should assessment look like, and how do we go about creditably assessing it? This paper addresses these questions with a focused inquiry of visual art assessment literature. The results strengthen the conceptual foundation for the “bundled” assessment model, confirm that authentic and performance-based visual arts assessment is greatly needed, and make clear that new tests should measure cognitive contributions found specifically in visual arts learning. The results also lead to the development and construction of sample items that can be “bundled” together by their type and kind, featuring a degree of item selection flexibility, using varied separate and aggregated scoring options, and employing a number of contemporary and folk art exemplars. This paper and the sample test items will significantly contribute to the body of knowledge about visual art assessment and give us a glimpse into what students learn specifically when they make, respond to, and think about visual art.

Developing A Bundled Visual Arts Assessment Model

Why is assessment important in art education? In a presentation on assessment at the 2005 Florida Art Education Association Conference, in a tongue-in-cheek manner I asked, What should visual arts assessment look like? a bread box? a NAEP item? Siege-smund item? a Dorn portfolio? an art history test? a Florida Comprehensive Assessment Art Test? an Advance Placement portfolio? a state music test? or a conglomeration of existing district art education test items?

The idea of how to credibly assess learning in visual art is elusive and confusing. All art teachers do some sort of assessment, but like art making itself, very few do it the same way or with much consensus. Armstrong (1994) provides two crucial reasons why it is important to develop and construct sound assessment practices. The first reason is to determine whether what we think is being taught is in fact being learned, and the second is that assessment results can demonstrate the value of art education in terms that the public can understand.

It seems that the emphasis on assessment in art education has ebbed and flowed, mostly around the formation and administration of the fairly controversial National Assessment of Educational Progress (NAEP) Arts Report Card (Persky, Sandine, & Askew, 1998). NAEP test items engaged students in creating, performing, and responding via constructed response items requiring students to write, draw, and work with collage materials. It has also been difficult to determine exactly what the role of national and state art education associations has been in developing assessment research and design. In the state of Florida, there has been interest in developing a test, similar to what the state music association has in development, but no specific plan is currently in place.

This inquiry began as a focused investigation of art assessment literature that includes theoretical foundations and recommended assessment testing formats. Of most interest was finding actual art assessment items. The task was then to evaluate these items and compose new ones covering unaddressed learning areas. The review of literature and research in the field of assessment became the foundation for articulating, defining, and creating parameters for the construction of a “bundled” assessment model. This new
bundled model (see Appendix A) could be tested and used in K-12 classrooms and potentially as an approach for state-level art assessment.

The conceptual strategy for a bundled approach refers to the type and kind of assessment measures to be used, varied aggregated scoring options, item selection flexibility, and the use of contemporary and folk art exemplars. It is important to develop separate and discrete measures to assess art knowledge, critical and aesthetic responses, and creating, rather than the combined constructed response blocks used by NAEP designers (Persky, Sandine, & Askew, 1998). Indeed, in order to more reliably assess art learning in these areas, it would be beneficial to rate each type of achievement discretely on separate item types and measures. Having the ability to assess students on discrete tasks and also on a bundled composite weighted rating would provide a much richer and more in-depth view of student art learning. Pistone (2002) gave a few examples of these types of “discrete” measures when developing drawing performance tasks that were not intertwined with short answer items or critical writing responses. The teachers and researchers could not only choose bundling according to the type and content of the items, they also have the option of combing various items, when, and how they see fit, on any given assessment. They might want to use a creating/drawing item and a critical/reflective writing item when a longer time frame is available for assessment, or bundle the multiple choice and matching items together when assessment time is limited. Another significant feature of this bundled approach is the interest in using contemporary and folk art exemplars and those related social and political issues that engage students more directly with the ideas and content found in their world today.

Review of Literature

The following focused review of literature was compiled to inform the foundation, development, and construction of the conceptual strategy for this assessment approach. The goal was to distill bundled multiple comprehensive items that could be used or modified in various K-12 settings at the district or state level. In the review, the goals were to look for quality of task or dependent variables, the validity of specific tasks, the content of task and item types, art exemplars, assessment measures, the role of student reflective writing, and demographic items.

In the course of this examination a variety of philosophical and ideological theories emerged about why art assessment is needed. We also discovered relevant ideas on how assessment should be conducted. Included are a variety of criterion and procedures that were tied to these philosophies. The last section includes examples of actual assessment items, and concludes by addressing shortcomings in the assessment literature.

Theory

Eisner (1996) begins by giving a number of accountability justifications for assessment and offers some criteria as guidelines. He says the failure to assess art education in ways that show distinct features of art learning may be a form of professional dereliction. Another purpose for assessment is to protect students from incompetent teaching and educationally diluted curricula. Without assessment in art education, there is no way to determine whether pedagogical practices have consequences and there is no basis for changing them. Eisner (1996) gives criteria he believes are important for assessment: Procedures should be user friendly; art tasks should be stimulating and relate to bigger ideas and should have multiple solutions; art programs should be congruent in content and aims; and content should address a variety of sensory and cognitive modes.

The assessment procedures suggested by Boughton (1996a) are based on criteria and protocol of the International Baccalaureate (IB). Boughton points to the widespread acceptance and the development of IB’s centralized assessment procedures.
He seems to reluctantly concede that there may be some value in developing at least a degree of comparability in assessment practices so that students' achievements may be tracked and understood on basic levels. His recommendation is to use a "processfolio" (à la Howard Gardner, 1996) that represents a kind of evolving cognitive map of work in progress. When it works well, the processfolio documents the student's growth and tracks the instructive value of all works. In a sort of economics-type disclaimer, Boughton points out that it is not realistic to expect this type of assessment to develop anytime soon, given the scarcity of funding, but he feels that such an assessment is valuable to art education.

Zimmerman (2003) suggests that art content and knowledge for assessment should be reflective of local community concerns and the needs of each teacher in his or her own art classroom. She adds that individual students' achievements should be measured against their own past achievements rather than solely against traditional standardized norms or criteria. Zimmerman, like many art educators, stressed using socio-anthropological bases for studying art works from a variety of cultures. Such study focuses on the sociocultural contexts in which works of art are created and stresses knowledge about the people who created them. It also includes folk and environmental arts from many cultures as well as traditional Western art.

Beattie (1997) responds to Boughton's NAEA Studies lecture in San Francisco (1996b) regarding problems of local versus national criteria, objectives, and standards for art used in visual art assessment. Both methods have helped to educate art teachers in ways that have clearly benefited teaching and learning. Beattie (1997) asks and answers the question in frequently expressed opposition: Should we have national standards for judging student performance in art education? The task of setting national standards and its probable assessment outcomes are too far removed from the context of the classroom. Beattie (1997) and Boughton (1996b) agree that assessment is meaningless without interpretation in a cultural setting; hence, assessment standards are best left to local school districts, schools, and teachers and students. The "bundled visual arts assessment model," as developed, provides for an opportunity to compare results among groups of students in the same school, different schools, and countywide comparisons. Further, as the NAEP (Persky, Sandine, & Askew, 1998) only provides for testing of one grade level, the bundled visual art assessment model can be adapted for use in a variety of school settings and in various grade levels.

Suggestions

Armstrong (1994) gives some very important suggested examples of different types of assessment instruments, as well as procedures for development, administration, scoring, and reporting results. She provides concrete examples (e.g., matching, alternative response, multiple choice, completion, short answer, and essay), all of which can be used for standardized assessment.

Dorn (2003) considered four research questions: Could the portfolio assessment process systematically quantify student art performances; was there inter-rater reliability among the teachers scoring the pre- and post-test portfolios scored as a combined group; were the raters' scores within each class normally distributed and did they provide sufficient score spread; and were gains or losses in student portfolio scores evenly distributed among students in the lower and higher performance categories? This study confirms that teachers who participate in training in portfolio assessment and curriculum development can conduct a valid and reliable assessment of student artwork. Dorn's type of authentic assessment calls for authentic performances, which include the behaviors of aestheticians, architects, art historians and critics, artists such as folk artists, people working in all forms who confront art in their daily lives, and people whose vocational activities relate to art. In regard to developing assessment models
he asks: How do assessment efforts relate to what teachers teach, what is the connection between school arts assessment and the school art curriculum, and how do we create an authentic assessment model that involves art teachers as stakeholders in the process?

McCollister (2002) offers ideas for writing critical essay assignments. The students’ writing may include background research, site visits, selection and use of an appropriate critical model, drafting the essay, peer editing, and final copy of edited writing. This is evaluated on criteria such as careful observation skills, the effective selection and use of a critical model, cohesive writing, interpretive content and meaning clarity, richness of language, and correct grammar use. McCollister presents ideas for the development of classroom rubrics that provide a description of the varying levels of productive completion such as novice, intermediate, and advanced. She suggests that first, one should decide what it is students needed to know while making the artistic process the primary goal based on national standards. Second, one should develop the assessment instrument on a topology of practice rather than an estimate of behaviors connected to selected art world figures. She may be alluding to one of the major criticisms of the NAEP (Persky, Sandine, & Askew, 1998), that the test did not reflect current practices and was too dependent on traditional and modern selected exemplary artworks.

Applications

Bezrucko’s (1995) study was driven by the assumption that visual arts teachers do not rely on objective achievement tests to assess learning and that teacher assessments of student art ability generally rely on subjective appraisal. Thirty-nine multiple-choice items were constructed to assess visual arts learning on 28 visual arts objectives in an Illinois State Board of Education statewide model fine arts curriculum. The test model for these items was a hierarchy of six internal components: Terms, tools, techniques, interpretation of an artist’s work, perceptual sensitivity to subtleties in visual artwork, and capacity to form cognitive inferences from visual information given to children in three instructional sections. The results indicated that the art-educated group scored significantly higher on the total test than did the non-art-educated group at all grade levels. Training and achievement showed a significant interaction in grade 7, indicating that additional years of art education yielded significantly higher visual arts achievement scores.

Howard Gardner (1996) describes Harvard Project Zero’s position that the trio of production, perception, and reflection is central to all art forms. Project Zero’s approach to assessment and curriculum was centered on the creation, and the development and pilot testing of domain projects and processfolios. In the “Composition” domain project, students develop an awareness of the basic principles of design. They create designs by randomly dropping and by deliberately arranging geometric shapes against a background. They then compare their patterns and reflect about the different effects achieved. Students were later introduced to certain major principles of composition (e.g., harmony through repetition, surprise through contrast), and discussed these principles as they were exemplified in paintings by professional artists. Finally, students create a composition in which they seek to achieve either harmony or surprise through one or more of the principles. Reflecting on their work, students evaluate the strengths and weaknesses of each one. They also use a “Biography of a Work” in which students first observe a large set of sketches prepared by Andrew Wyeth (prior to his completion of Brown Swiss) and then survey a companion set of sketches and drafts of Picasso’s Guernica. Following these perceptual explorations, they embark on creating their own paintings and drawings, and monitor their own “developing” processes. The students’ task was to make a picture of their room at home, bringing out aspects of their own personality in the way that they portray the room. This domain project features constant dialect between sketches and final products, and between
the preparatory works of major artists and one’s own “rehearsals.” The primary criticism of this example is the sequence of the drawing task and the obvious omission of contemporary artwork.

The driving purpose for the NAEP (Persky, Sandine, & Askew, 1998) assessment was based on the premise that all students should receive a solid arts education and that there is a great need for valid and reliable information on art learning particularly since the last National Art Assessment (National Center for Educational Statistics, 1974-1979) was conducted a generation earlier. The 1998 NAEP was considered an innovative demonstration assessment intended to reflect a comprehensive view of good arts education practices and the complex nature of art making. Some of the key foundations for the NAEP 1998 were that arts integrate intellect, emotion, and skills in making meaning and that arts education meets the demands of the workplace and develops reflective skills. Test items involved students in creating, performing, and responding tasks. Authentic tasks included constructed response items requiring students to write paragraphs through a process of analysis, description, and identification, after which they would draw and work with collage materials. Demographics and information about instruction and items that examined self-perception about art abilities were also included. In the procedures, students were introduced to works and creating activities with text and visual examples (exemplars). Carefully crafted instructions guided students through sequential step-by-step tasks to give as much information as possible about what they would demonstrate. Responding items were built around a theme or concept for students to focus on in a timed assessment. One of the largest challenges was scoring student artwork and in training raters to apply criteria. Raters of 3-D works judged student artwork from photos; and the NAEP researchers warned that such ratings may be problematic.

Seigesmund, Diket, and McCulloch (2001) reported that the NAEP (Persky, Sandine, & Askew, 1998) results were not particularly encouraging. Art instruction seemed to have negligible influence on outcomes because only 6% of the students received “effective/adequate” scores on creative production, and only 48% reported they were currently enrolled in art classes. The researchers asked: What’s going on here? In the 1998 NAEP “Collage Block,” students were asked to respond to 10 questions then make their own artwork. The problem that Seigesmund, Diket, and McCulloch reported with the Educational Testing Service administration was that students had not done this type of art-making procedure before. It had also been assumed in the prior administration that a visual prompting of exemplary artist’s work was an often-used curricular format to generate student art production. Seigesmund, Diket, and McCulloch decided to use the same school site, to reconfigure and rescoring, and to use new responding, exploring, attending, relating, and sustaining criteria to evaluate student work. By re-administering the NAEP “Collage Block,” the authors now had some baseline data along with student reflections about their art making that have always been considered a part of a comprehensive art education.

Seigesmund, Diket, and McCulloch (2001) readministered the test item and some of the key findings included: Reconfigured scoring increased validity and reliability, teachers were better able to see students’ individual abilities, the test exposed deficiencies in curriculum and state standards (inadvertently), the new rubric tightened the gap for socioeconomic variances, and the school received accurate quantitative data to start tracking student performance in visual art. Although findings indicate that there is no predictive achievement in non-art subjects, this new test measured a cognitive contribution not made in other areas such as the ability to recognize and employ construction and meaning, and that achievement in visual arts does not necessarily explain success in other subjects.

Dorn’s (2003) project assessment activities included training teachers in the use of art rubrics in assessing pre-K-12 student art
performance, using blind scoring methods by peer teachers to validate teacher-scored student work, training in the use of authentically scored student art as a curriculum tool for the improvement of art instruction, developing assessment portfolios and analytical rubrics for special needs, and developing assessment instruments and methods of reporting consistent with student needs and with state and school district standards. In-service institute instructors included artists, curriculum and assessment specialists, and art educators. This instruction was used to ensure the philosophical validity of the teachers’ curriculum, which assured consistency with the means and ends of art, and provided for accurate and significant representation of the products of artistic inquiry. The curriculum and assessment specialists assisted the teachers in writing lesson plans, developing rubrics and portfolios, and creating methods for reporting the results of assessments. Dorn pointed out that the most important concern in the physical design of the performance assessment was that it reflect the nature of the exercises already embedded in the art curriculum and that it encourage students to study their own trains of thought as revealed in notes, sketches, and practice efforts. He found that there are viable alternatives to multiple choice, matching, and short answer tests in art assessment and that all art teachers not be expected to teach, nor all students need perform, in the same way.

Dorn (2003) also gives some valuable insights for further study that are applicable to this research. He suggests using various rubrics, teacher-constructed tests measuring content knowledge, and approaches to student self-evaluation. Such varied measures could strengthen the predictability and better assess what additional forms of art classroom performance need to be considered. Dorn feels an attempt should be made to develop a system-wide assessment plan using some of the procedures used in his study.

According to Yan and Rieder (2001), as of October 2001, nine states (Illinois, Kentucky, Maine, Minnesota, Missouri, New Jersey, New York, Oklahoma, and Washington) had mandated requirements for fine arts assessment and seven states (Alaska, California, Massachusetts, Ohio, Pennsylvania, Utah, and Vermont) had voluntary fine arts assessment. Only Kentucky and Minnesota have high-stakes (used in school accountability or student graduation requirements) arts assessment. Not every state assesses at all grade levels (K-12), but most assess in music, dance, theatre/drama, and visual arts. The most common and most often used item formats are multiple-choice and constructed-response items. Even though many states have a stand-alone arts assessment, the number of items is very limited and can range from as few as 10 to as many as 50. This view of statewide large-scale fine arts assessment demonstrates a great need for this paper and for a study to address the need for more and better visual arts assessment instruments.

What Should it Look Like?

So what did we learn about developing and constructing a bundled visual art assessment from this focused review of literature? What was gleaned from the investigation and what should this assessment model look like?

All of the authors cited in this review affirmed that there is a lack of, and great need for, visual arts assessment at local, state, and national levels and that these assessments should be more than multiple choice, matching, and short answer tests. There was also a consensus that the NAEP Arts Report Card (Persky, Sandine, & Askew, 1998) results were not encouraging: As Seigemund, Diket, and McCulloch (2001) suggest, there is a need for new tests to measure cognitive contribution made specifically in visual arts and that are not accounted for or measured in any other disciplines.

Based on Pistone’s (2002) assessment examples, bundled assessment items should be authentic and performance based, in that they should include student artwork and what the students demonstrate they learn from making art. It was also found that there are considerable curricular concerns
and issues about the art tasks used in some research studies (Gardner, 1996), and there is an interest in the study of more contemporary art and folk art exemplars (Zimmerman, 2003). There is a pronounced need for more critical and reflective writing, not only about student artwork but also about works made by contemporary artists (McCollister, 2002).

The sample bundled assessment items that follow (see Appendix A) fold in the recommendations from the review of literature into discrete measures of what selected 8th-grade students might know and be able to do. It is expected that the results of these items will be a more accurate and authentic measure of student art learning.

Conclusion

As art educators, we should be past the point of being afraid of, or threatened by assessment. As a discipline we need it, but two persistent questions remain; what should assessment look like, and how do we go about creditably assessing visual arts learning? To address these questions, this paper began with a focused inquiry of visual art assessment literature to help build and enhance the conceptual framework for a bundled visual arts assessment model and to guide its development and construction. The investigation examined the content and quality of dependent variables (actual art assessment items or tasks) as well as the type and kind of items. The examination also evaluated the type of art exemplars recommended and used in the literature, and the role student critical and reflective writing played in these discussions.

The results are presented in two parts. The first part enhanced and strengthened the conceptual foundation for the assessment model. It appears to me that there is a general consensus that visual arts assessment is greatly needed at all levels and it should include more than multiple choice, matching, and fill-in-the-blank type test items. There was also a sentiment that the NAEP Arts Report Card (Persky, Sandine, & Askew, 1998) results were discouraging and new tests to measure cognitive contributions specific to visual arts were needed. I agree that there should be more authentic performance assessment that includes student artwork, and what students demonstrate they learn from making art. Curricular questions about the art tasks used in some studies were raised along with the lack of contemporary art and folk art exemplars. More critical and reflective writing was recommended.

The second result impacted the development, construction, and presentation of a few sample bundled assessment items (see Appendix A) built on the strengths of previous assessment research. The sample items can be “bundled” together by the type and kind of item, featuring a degree of item selection flexibility, using varied separate and aggregated scoring options, and employing a number of contemporary and folk art exemplars. This paper and the sample test items will significantly contribute to the body of knowledge about visual art assessment and give us a glimpse into what 8th-grade students learn specifically when they make, respond to, and think about art.

Author’s Notes: I would also like to acknowledge the work of a number of people in the preparation of this manuscript: Kathleen Shannon, Deborah Umphrey, Arna Watson-Young, Hope Conner, Anne Grey, Carolyn Ibarra, Kathryn and Virginia Farmer, Dr. Chuck Dziuban, and Dr. Patsy Moskal. This paper and line of research is presented in the memory of my major professor Dr. Charles M. Dorn who first introduced me to the concept of art assessment and devoted such a large part of his professional life to improving the assessment of teaching and learning in our field.

This research is sponsored in part by a 2007 National Art Education Foundation Research Grant and the University of Central Florida.

References


Correspondence concerning this article should be sent to Professor Thomas M. Brewer, Coordinator of Art Education, Department of Teaching and Learning Principles, College of Education, University of Central Florida, Orlando, FL 32816-1250, phone 407-823-3714, email: tbrewer@mail.ucf.edu
Appendix A

Sample Bundled Test Items

Multiple Choice Sample Items

Directions: Read each statement and review all possible answers. Circle the letter of the correct answer.

1. A(n) ___________ gets and displays collections of art works to educate the public.
   a. Curator
   b. Graphic artist
   c. Illustrator
   d. Sculptor

2. Look at contemporary artwork Man Spirit Mask, 1999, by Willie Cole (below). Through the use of symbolism Cole is using the shape of an iron to communicate what social message to the viewer? Circle the letter of the correct answer.

   a. ancient symbol used on the staffs of African leaders to show their power
   b. metaphor for a slave ship used to bring slaves to America
   c. image engraved on their shields during tribal warfare to identify their tribe
   d. symbol for the struggles of his mother and grandmother’s life as a domestic worker

3. Alice Neel’s Mother and Child, 1967 (right) would be a(n) ________________ example of art.
   a. Abstract
   b. Expressionist
   c. Fauvist
   d. Pointillist

a. specific historical event  
b. political imagery to impact people  
c. family event to remember the ancestors  
d. mythological event that is told by storytellers

5. Judging from the reproduction on the right, *Cotton Pickin’ Time*, a painting by Clementine Hunter, what answer **best** describes the idea the artist is trying to communicate to the viewer?

a. shows that many Americans labored in agriculture  
b. shows social life, people and surroundings  
c. shows people engaged in leisure activities  
d. captures a specific event in history

Matching Sample Item

**Directions:** Each image below represents one art theory. Based on your own judgment select a theory below and place the letter of that theory on the blank line under the image.

**Art Theories:** A= Expressionism; B= Instrumentalism; C= Formalism.

Jose Angel Toirac, *Obsession*, 1996

6. _____________  
7. _____________  
8. _____________
Art Analysis Writing
9. **Directions:** Some artists express their views on particular issues in their paintings. Look at Frank Moore’s *Black Pillow, 2002* (below). In your own words, **describe** the subject matter, **analyze** the formal qualities, **interpret** the meaning, and **judge** (Feldman, 1994) the overall success of the painting.

Drawing Task
10. **Directions:** For the next 45 minutes, you will make one small planning sketch and one final larger, more finished 10” x 8” drawing. First, you will be given 15 minutes to make one planning sketch. Together we will slowly read the instructions twice to be sure you understand the possible subject matter and meaning you will develop in the sketch. Please take notes, mark highlights, and list your ideas on the instruction sheet. Second, you will have 30 minutes to complete your final finished drawing based on your sketch and notes. Here are the instructions:

“As you have seen from the contemporary artwork on this test, artists often communicate through social observation or commentary in our everyday lives and contemporary artwork. As artists, you must also be reporters and interpret the role of culture and society and communicate that message to others. Create a drawing that you as an artist will use to report to others. Suggestions range from something you have seen on TV news, or you may prefer some meaningful activities in your everyday life (families participating in domestic activities, celebrations, or work-related activities). Just as in reporting the news, art can praise the good (people helping people) and expose threats in society (hunger or homelessness).”
Figure/Item Image List


Figure 3/Item Image 4: Pettibon, Raymond (Artist). (2004). *No title (there will be)* [Ink and Watercolor on Paper, signed and dated on verso], Santa Monica, California, Ikon Ltd., Kay Richards Contemporary Art.

Figure 4/Item Image 5: Hunter, Clementine (Artist). *Cotton Pickin’ Time* [Painting], Baton Rouge, Louisiana, Louisiana Division of the Arts.


Figure 8/Item Image 9: Moore, Frank (Artist). (2002). *Black Pillow* [Oil on Canvas over Featherboard Panel with Gouache on Paper Attachments], Orlando, Florida, Collection of the Orlando Museum of Art, Gift of Council of 101, 85.16