DOCUMENT RESUME

ED 370 957 TM 021 098

AUTHOR Bezruczko, Nikolaus
TITLE Development and Evaluation of a Visual Arts Achievement Test.
INSTITUTION Chicago Board of Education, Ill.
SPONS AGENCY Illinois State Board of Education, Springfield.
PUB DATE Apr 92
NOTE 46p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Art Education; Content Validity; Correlation; Educational Assessment; Elementary School Students; Elementary Secondary Education; Evaluation Methods; Grade 3; Grade 7; High School Students; *Item Response Theory; Kindergarten; Reliability; Scores; *Test Construction; Test Items; Test Validity; *Visual Arts
IDENTIFIERS Chicago Public Schools IL; Rasch Model

ABSTRACT

Internal structure and external validity of 39 multiple-choice visual arts achievement test items were examined. These items were developed to assess grade 3 visual arts achievement for a statewide model of a fine arts curriculum. Item responses were evaluated in terms of: (1) fit to the one-parameter Rasch measurement model; (2) item-total correlations and alpha reliability; (3) total score comparisons between art- and non-art-educated groups in kindergarten, grade 3, grade 7, and high school (over 900 students in all); and (4) comparison of art- and non-art-educated groups on six components of visual learning. Most items generally fit a unidimensional measurement model, with good alpha reliability, although six items showed marginal or poor fit. Art-educated students scored higher in each grade, and when items were grouped into the six components of visual arts achievement, art- and non-art-educated students differed significantly as expected, except for knowledge of tools, where no significant difference was noted. This method of assessment appears reliable and valid for children in grade 3 and may be useful for older children as well. An appendix contains a long table of curriculum objectives and interval structure. (Contains 26 references.) (SLD)

Development and Evaluation of a Visual Arts Achievement Test

by

Nikolaus Bezruczko

Chicago, Illinois

Running Head: Visual arts achievement test

Development and Evaluation of a Visual Arts Achievement Test

Abstract

This study examines the measurement properties, internal structure, and external validity of 39 multiple-choice visual arts achievement test items. These items were developed to assess grade 3 visual arts achievement on a statewide model fine arts curriculum.

Item responses were evaluated in terms of a) fit to the one-parameter Rasch measurement model; b) item-total correlations and alpha reliability; c) total score comparisons between art- and non-art-educated groups in kindergarten, grade 3, grade 7, and high school; and d) comparison of art- and non-art-educated groups on six components of visual arts learning.

The results showed that a test of 39 items for the overall group generally fit a unidimensional measurement model, and the alpha reliability was good (R_{tt} = .86). Six items, however, showed marginal or poor fit. The reliability of 38 items for grade 3 was comparable (R_{tt} = .81), and one item showed marginal fit.

Comparisons between art-educated and non-art-educated students in the overall group showed the art-educated students to score significantly higher in each grade with a significant interaction between years of art-education and total test scores in grade 7.

When 39 items were grouped into six components of visual arts achievement, art- and non-art-educated students differed significantly in the expected direction on all components except knowledge of tools. In grade 7, art and non-art-educated students did not show a significant difference in their knowledge of visual arts tools.

In conclusion, this method of assessment appears reliable and especially valid for children in grade 3 and may provide insight into the visual arts achievement of older children as well.

Development and Evaluation of a Visual Arts Achievement Test

Unlike teachers of basic school subjects such as reading or arithmetic, art educators do not rely on objective achievement tests to assess visual arts learning. They tend not to view test items as valid sources of information concerning visual arts learning or performance on tests as an appropriate goal of visual arts education. Art educators emphasize the personal interpretation of artistic experience and the ability to critique aspects of artistic productions more than the acquisition of objective knowledge (Eisner, 1985). Likewise, teacher assessments of student art ability generally rely on subjective appraisal.

Evaluators of school programs and some classroom art teachers, however, emphasize their need for objective evaluations of visual arts learning. Hoepfner (1984), for example, described testing and measurement procedures needed to evaluate visual arts programs, and Frechtling (1991) noted the desirability of using traditional standardized testing methods that *complement* performance-based assessments. Art education reformers have encouraged research into the reliability and validity of standardized visual arts achievement tests (Getty, 1985).

Although not explicitly stated, cognitive researchers imply their need for objective assessment methods when they speculate on fundamental relations between visual learning and cognitive development. Gardner (1982, 1983), for example, suggests that visually manipulating symbols during art production underlies the process of language acquisition, as well as complex mental thought. Other researchers, (Arnheim, 1969, 1986; Ecker, 1963; for a review see

Hamblen 1992) speculate that art experience promotes problem solving and thus influences intellectual development. Consequently, objective methods to measure visual arts achievement should improve empirical investigations into these relationships.

Purpose

The purpose of this study is to conduct a rigorous evaluation of the measurement properties, internal structure, and external validity of 39 multiple-choice test items designed to assess visual arts learning in grade 3. These items are unusual because, first, they assess an area of school learning virtually untouched by modern measurement technology and, second, the items assess a wide range of visual arts learning from simple knowledge of art-related terms to complex perceptual and cognitive judgments.

This evaluation addresses the following issues.

1. Do child responses to multiple-choice visual arts test items have psychometric properties of reliability and internal consistency? A related concern are the visual characteristics that distinguish between difficult and easy items.

2. Do visual arts test items have criterion validity? In particular, do art-educated children receive higher scores? Likewise, does years of art education show relations to test scores?

3. Do visual arts test items show construct validity? Are differences in art background significantly related to scores on test

components of visual arts achievement (i.e., knowledge of terms, tools, techniques and so on), and are the difficulties of the respective test components, relative to each other, theoretically plausible?

4. Do art assessment items that rely on high quality photographic reproductions validly assess children's awareness of qualitative characteristics such as texture, color, movement, and their interrelations? Do these items assess children's understanding of the artistic process?

Review of Standardized Visual Arts Achievement Tests

Despite an apparent need for objective methods, little refinement or adaptation of contemporary objective testing methods to the visual arts has been undertaken. The only art achievement test, for example, in the Ninth Mental Measurement Yearbook (Mitchell, 1985) is the NTE Specialty Area Test in art education for college seniors and teachers.

Nationally, assessments of art knowledge and attitude have been conducted by the National Assessment of Educational Progress (1978a, 1978b, 1981). Their assessments and analyses, however, were not intended to advance an understanding of valid or reliable visual arts assessment or related issues concerning the dimensions of learning that underlie art achievement. Consequently, their results do not provide insights into methods that are appropriate for measuring visual arts learning.

An effort is currently underway in Minnesota (Higgins, 1989) to implement a statewide plan of visual arts assessment. The method involves the development of a centralized item bank of multiple-choice

visual arts test items, but information concerning its success is not yet available.

In the United Kingdom, Bennett (1989) described the assessment of children working towards the General Certificate of Secondary Education in art and design, a method that relies on curriculum objectives and criterion referenced test items. Although he provides some insight into the adaptations that are needed to apply traditional testing methods to the visual arts, he prefers alternative assessment methods.

The Problems

Objectively assessing visual arts learning presents test developers with several problems concerning a) test items that present written content, b) items based on poorly defined factors of visual arts learning, and c) reliability and validity.

First, although written test items commonly assess knowledge of art history and design principles, this method is troublesome when assessing visual arts learning. This approach is especially inappropriate for young children -- a special category of art student -- because it always runs the risk of primarily assessing reading achievement rather than visual arts learning. Attempts to alleviate this problem by developing test items with photographs, however, have been limited by the unavailability of appropriate artwork (Hoepfner, 1984), as well as technically inadequate photographic reproductions. Bennett (1989) especially objected to photographic representations of complex texture, color, and form inter-relations in original artwork.

Second, art educators and curriculum developers do not agree on the underlying factors of visual arts achievement. Although art curricula are usually based on implicit notions of learning such as a) knowledge of tools and art-related technical terms, b) perceptual sensitivity, c) visual cognition of thematic content, and so on, art teachers tend not to use these criteria systematically. Consequently, the vague relations between assessments based on explicit performance criteria and practical instructional goals make much contemporary art assessment meaningless.

Third, art experts and educators seriously object to applying the requirements of replication and standardization, the foundation of educational evaluation, to behavior commonly associated with originality and creativity such as artistic production. Objective assessment depends on comparing tangible evidence of child learning (i.e., responses to test items, products of student performance, or some alternative) to school expectations or standards which enable teachers and parents to form judgements concerning mental ability and acquired skills. This process, however, emphasizes learning that conforms to explicit and uniform performance criteria and thus undermines the influence of creativity on child performance.

Method

Sample

Selection of schools. Intact classrooms in kindergarten, grade 3, and grade 7 in ten elementary schools and two high schools of the Chicago Public Schools were selected to participate in this study. Four elementary schools and one high school with visual arts education were matched socio-economicly and academically with six elementary schools and one high school without visual arts education. (The difference in number of schools was necessary to ensure that art- and non-art-schools would be adequately represented.)

The five schools with visual arts programs (henceforth called the art-educated group) employ full-time art teachers with specialized education, and children begin their participation in the programs when they enroll in kindergarten. In interviews, principals and teachers in these schools emphasized the importance of visual arts education for children.

The schools without art education (henceforth called the non-art-educated group) do not employ trained art teachers, and the art experiences that these children receive is provided at the discretion and convenience of their classroom teachers. Teachers and principles in these schools tend to emphasize the learning of basic school skills (reading and arithmetic) and do not emphasize visual arts experiences. Consequently, visual arts education for these children varies from year to year depending on available resources and teachers' personal interest. Table 1 presents a description of the schools.

Table 1: Description of the Schools

School characteristics	Elementary schools										High schools	
	A			B		C			D		E	
	Art[A-1]	Nonart[A-1]	Nonart[A-2]	Art[B-1]	Nonart[B-1]	Art[C-1]	Nonart[C-1]	Nonart[C-2]	Art[D-1]	Nonart[D-i]	Art[E]	Nonart[E]
Stability	97.8	97.0	98.7	97.4	95.1	81.9	90.4	91.9	89.2	85.9	76.0	79.0
Attendance	95.5	94.1	91.3	94.6	93.6	93.8	94.0	91.3	93.0	92.1	83.0	87.4
Low income	40.2	35.5	49.6	5.4	4.6	75.2	52.2	61.3	12.4	58.4	38.6	3.6
Limited English	10.4	7.8	0.0	0.0	0.0	0.0	0.0	0.0	1.0	22.7	1.3	1.8
White nonminority	30.0	26.4	21.5	57.0	61.0	0.0	0.2	0.3	41.1	0.2	26.3	18.3
African-American	40.0	30.6	78.1	33.0	37.0	93.1	99.3	99.1	42.8	60.3	49.2	63.3
Other minority	30.0	43.0	0.3	10.0	2.0	6.9	0.5	0.2	15.2	39.6	24.5	18.5
Neighborhood code[1]	P	A	P	W	W	P	P	P	P	P	U	W
N =	88	60	55	88	94	79	87	59	55	53	230	83

Note. All values are percents. Numbers may not add to zero because of rounding. The information in this table was provided by the Chicago Board of Education (1989, 1991).

Student characteristics. On the whole, the sample represents several ethnic and racial minorities, although in two schools, white nonminority children were the majority. Overall, 48% of the art-educated children were girls and 52% boys. In the non-art-educated group, 44% were girls and 56% were boys. Students generally reflected the socio-economic characteristics of their respective schools as they are presented in Table 1.

Visual Arts Assessment Items

This evaluation concerns 39 multiple-choice test items that assess children's performance on 28 visual arts learning objectives in a statewide model fine arts curriculum (see Bezruczko, 1989). A brief description of the learning objectives appears in the Appendix.

These objectives, and the items that assess learning on them, are related to six general areas of visual arts instruction presented below. Because knowledge in these areas contributes to the overall visual learning of children, in this study they are called components of visual arts learning.

- Knowledge of terms
- Knowledge of tools
- Knowledge of techniques
- Interpretation of an artist's affective intent

- Perceptual sensitivity to subtleties in an artwork

- Capacity to form cognitive inferences solely on the basis of visual information.

Knowledge of terms, tools, and techniques. These components assess knowledge of specific tools used in the production of visual art, the correct use of terms associated with the production of artwork, and the technical process of forming raw materials into finished artwork. Teachers expect this knowledge to be necessary for children to advance to higher levels of artistic knowledge and to enhance their general appreciation of visual art.

Interpretation of an artist's affective intent. Although a viewer can never really know an artist's intention solely through an artwork, test items can assess children's ability to relate physical characteristics of an artwork to its affective response and thus infer a reasonable intention. Teachers, for example, expect children to understand that the smile on a figurine or in a painting was probably intended to convey some aspect of happiness, and that it is an objective characteristic of the artwork. While teachers and children may differ in their interpretation of its significance, the artist's intention here becomes an important aspect of the finished work.

Perceptual sensitivity to subtleties in an artwork. Art teachers and laypersons commonly believe that art education promotes perceptual appreciation of visual art. Art teachers emphasize sensitivity to the fine detail of line quality, variation in the shading of color, and the interplay of image and space in an artwork, and an effective assessment of visual

learning will show the importance of this component to children's visual arts achievement.

Capacity to form cognitive deductions. Experts and laypersons alike are aware that art is characterized not only by beauty, but by thematic content that is to some extent independent of one's appreciation of the artwork. The intellectual ability to separate thematic content from physical beauty in artwork several centuries old, and mentally manipulate this information, can provide children with powerful insights into the influences that shape civilization and contemporary life. While this component represents a complex goal of art education, art teachers try to teach thematic understanding, and thus it should be represented in an assessment of visual learning.

Ordered relations of the components. Because assessment items differ in difficulty, some requiring simple and others more complex knowledge, child responses to items establish levels of achievement that are interpretable to teachers. In theory children who only pass items that assess knowledge of terms are showing a lower level of visual arts achievement than children who pass items that require visual information to make complex cognitive deductions.

This consideration of components is necessarily speculative. An empirical analysis is needed to precisely order the components that are associated with visual arts learning, identify the reliability of an ordering, as well as estimate the magnitude of difference between components for art-educated and non-art-educated students.

Expert reviews of the items. Concurrent validity between item content and learning objectives was established by a panel of twelve reviewers consisting of art teachers, museum specialists, and curriculum evaluators. Only items with 100% agreement were recommended for

assessing visual arts learning.

Production of the items. Reading vocabulary of the items was controlled to not exceed the third grade. Each item used either full color or black and white photographic reproductions of authentic artwork to assess learning of a particular curriculum objective (see Appendix). The items were reproduced on card stock and bound into a booklet (Bezruczko, 1989).

Research Plan

In order to conduct this study, the following activities were completed.

- Visual arts test items were constructed to assess learning on specific objectives in a statewide model visual arts curriculum (Illinois State Board of Education, n.d.).

- Visual arts test items were administered to children in kindergarten, grade 3, grade 7, and H.S. in schools with and without visual arts education.

- Item responses were statistically analyzed for fit to the Rasch measurement model, alpha reliability, and validity.

- Groups of items were identified defining six components of visual arts achievement (i.e., terms, tools, techniques, perceptual sensitivity, interpretation of affective intent, and cognitive deductions) and their scores were examined across

grades 3, 7, and high school.

Procedures

The item booklets were administered to intact classes in kindergarten, grade 3, grade 7, and high school. The questions were read to children in kindergarten and grade 3. All other children read the items to themselves. The children in grade 3 and above marked their answers on an answer sheet. All items were administered in a single session and none of the children were prevented from completing the booklet because of time.

Analyses

Empirical analyses were conducted to establish the measurement properties of the items in addition to internal structure, and criterion validity of a test based on them. Consequently, analyses were conducted of the overall sample, then only grade 3 -- the target population of the assessment items.

Measurement properties and internal structure. Item difficulties and model fit *t* values were estimated using the one-parameter Rasch measurement model (Wright & Linacre, 1992) of the overall group and grade 3 children. These analyses were supplemented with an examination of item-total correlations and alpha reliability. In addition, a principle components factor analysis was conducted of the item responses by the overall group.

Analyses of item difficulty and fit to a linear measurement model were conducted to establish the measurement properties of child responses to visual arts assessment items. The factor analysis and item-total correlations were conducted to provide a description of the internal structure of these items as a standardized test.

Validity. Criterion validity was established by examining a 2 X 4 analysis of variance of the total test scores between art-educated and non-art-educated children and between grade levels. A valid test of visual arts training will show that art-educated students score consistently higher on the assessment items than non-art-educated children.

Construct validity was investigated by comparing the obtained ordering of internal components with theoretical expectations. Because art experts and educators theoretically consider knowledge recall a lower level mental process than cognitive reasoning or perceptual interpretation, they should be significantly easier to pass. Consequently, a 2 X 4 X 6 analysis of variance examined the performance of art-educated and non-art-educated children by grade level and by test component.

Results

Overall Group

The Appendix shows the *p* values, item-total correlations, transformed item difficulties, and Rasch infit *t* values of the items for the overall sample.

Measurement properties. Rasch infit *t* values, a statistical means

of identifying item response patterns inconsistent with linear measurement, were larger for Items 2, 5, 10, 20, 22, and 27 than the *t* criterion ($t = 3.00$) set in the measurement model (Wright and Stone, 1972). For these items, *t* values are 3.3, 4.3, 3.6, 3.5, 5.5, and 3.1, respectively ($N = 1,001$). All the misfitting items except Item 2 represent the component testing knowledge of terms, and they are relatively easy items ($p > .95$). Item 2 represents the component testing knowledge of tools.

This discrepancy between obtained and expected values by the Rasch measurement model is important for two reasons. First, positive misfit means that significantly more children with low scores on the total test of 39 items passed more relatively difficult items than mathematically predicted by the measurement model. Consequently, their numerical measures may not validly represent their qualitative ability. Second, the analysis of fit establishes the integrity of the test as a measuring procedure. If many items or many persons misfit the model, the test fails to function as a measuring process.

Internal structure. *P* values of the items ranged from .39 to .97 with a mean of .80. None of the items showed ceiling or floor effects. Items 8, 9, 11, 23, 28, and 30 were the easiest ($p > .90$) where Items 8, 9, 11, and 23 tested the component terms and Items 28 and 30 tested the component perceptual sensitivity. Items 22 and 40 were among the hardest ($p < .50$) and also tested knowledge of Terms. Items testing cognitive deductions were the hardest.

Item total correlations ranged from .18 to .48 with an average of .33. The alpha reliability of 39 items for the overall group was .86.

Factor analysis. A principle components factor analysis yielded four factors with eigenvalues greater than 1.0 that showed interpretable

content. The eigenvalues were 6.19, 2.14, 1.98, and 1.41 accounting for 15.9%, 5.5%, 5.1%, and 3.6% of the variance, respectively.

Factor 1 showed 13 items with positive loadings greater than .30 and none of the items showed negative loadings. These items test the ability to identify aspects of mood and emotions, as well as knowledge of techniques associated with the production of sensory effects in visual art.

Factor 2 showed 11 items with positive loadings greater than .30 and none with negative loadings. The content of the items test the ability to identify physical aspects of artworks such as texture or rhythm. Factor 3 showed 7 items with positive loadings greater than .39. All of these items assessed knowledge of drawing tools.

Factor 4 showed 6 items with positive loadings greater than .40. These items did not show interpretable content and all of them had large Rasch infit values ($t > 3.00$).

Third Graders

Table 2 presents the internal structure and measurement properties of the items for the third graders.

Measurement properties. None of the items in grade 3 exceeded the fit *t* criterion of the measurement model.

Internal structure. *P* values ranged from .23 to .97 with a mean of .72. None of the items showed ceiling or floor effects. Items 8, 9, 1[illegible] 14, and 18, were the easiest ($p > .97$), and except

Table 2: Internal Structure for the Third Graders

Item No.	P-value	item-total	Rasch[2] logit	fit t[3] value	Item No.	P-value	item total	Rasch logit	fit t value
2	.61	.28	.82	1.4	23	.89	.46	-1.05	-1.7
3	.82	.32	-.52	-.2	24	.63	.36	.78	-1.1
4	.83	.17	-.52	1.3	25	.82	.30	-.42	-.9
5	.45	.24	1.58	2.3	26	.56	.20	1.09	1.1
6	.39	.25	1.90	1.8	27	.46	.13	1.56	2.5
7	.92	.27	-1.51	.1	28	.89	.27	-1.09	-.8
8	.97	.28	-2.55	-.4	29	.78	.29	-.16	-.1
9	.96	.21	-2.31	-.5	30	.89	.19	-1.14	-.6
10	.65	.29	.57	1.9	31	.75	.17	.09	.8
11	.95	.28	-2.04	-.7	32	.82	.41	-.39	-2.1
12	.83	.29	-.57	-.3	33	.64	.42	.75	-1.6
13	.81	.35	-.32	-.6	34	.74	.38	.15	-.6
14	.93	.14	-1.73	-.1	35	.62	.46	.83	-2.1
15	.83	.41	-.53	-.8	36	.82	.19	-.43	-.1
17	.88	.28	-1.00	-.7	37	.45	.34	1.65	-.6
18	.93	.30	-1.60	-.4	38	.65	.26	.65	.5
19	.58	.25	1.00	1.7	40	.31	.24	2.36	-.2
20	.44	.31	1.62	1.0	41	.70	.38	.37	-1.4
21	.82	.29	-.44	-.3	42	.79	.40	-.19	-.6
22	.23	-.12	2.77	3.8					

Note: *N*s for the items ranged from 335 to 373.

for Item 14 all of them tested knowledge of Terms. Item 14 tested knowledge of techniques. Items 5, 20, 22, 27, 37, and 40 were the hardest items ($p < .50$) also representing the component Terms. Item total correlations ranged from -.12 to .46 with a mean of .28.

Alpha reliability for grade 3 ($N = 250$) based on 38 items was .81. (Item 22 was dropped because its item-total correlation was negative and fit t was large, $t = 2.77$.)

Validity

Analysis of variance of total test scores by grade and education. Table 3 presents the means and the standard deviations of the total test scores. The results of an analysis of variance in Table 4 and Figure 1 show the total test scores to differ significantly between grades and between art- and non-art-educated groups, and that the difference in magnitude increases after grade 7. The significant interaction in grade 7 means that children with the most education receiving the highest scores.

Analysis of test components. Table 5 presents the means and standard deviations of the test components for the art and non-art-educated students. An analysis of variance in Table 6 shows that the scores significantly increased for each grade and that art-educated children scored significantly higher on all components except knowledge of tools. Figure 2 shows the scores of the components after transforming them to one-parameter logit scale values.

These results suggest that children enrolled in art education, not surprisingly, learn more about the instructional content assessed by these items than non-art-children. The group differences,

Table 3: Means and Standard Deviations of the Total Test Scores

	Art		Nonart		
	Mean	SD	Mean	SD	N
K	9.78	10.49	8.62	8.48	26-27
Grade 3	29.66	4.64	27.74	5.49	104-147
Grade 7	33.16	4.01	32.30	4.04	113-151
High school	34.87	4.43	32.94	4.75	64-194

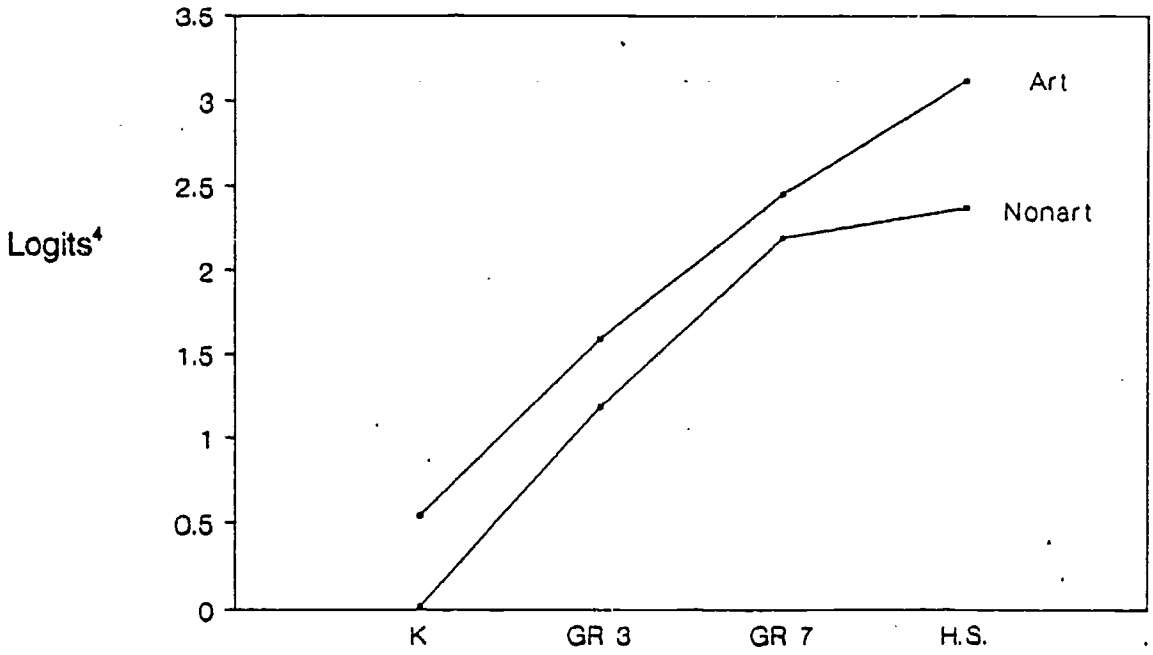
Table 4: Sum of Squares

Source of Variation	Sum of Squares	*DF*	Mean Square	*F*	*p*
Grade	320.41	2	160.20	114.90	<.001
Education	49.70	1	49.70	35.64	<.001
A X T	12.16	2	6.09	4.37	.01
Error	1328.79	953	1.39		

Note. Total scores were transformed to one-parameter logits. This comparison includes only grades 3, 7, and H.S. students.

Figure 1

Comparison of Art Achievement Test Scores



Note. The kindergarten comparison is based on 53 children. The *Ns* for the elementary grades range from 127 to 375 and the high schools, 313.

Table 5: Means and Standard Deviations of the Test Component Scores

Component scales	Background	Grades[5] 3	7	H.S.
Terms				
	Art	19.38	21.68	23.02
		3.26	2.96	2.97
	Nonart	17.89	21.09	21.72
		3.79	2.74	3.17
Tools				
	Art	7.14	7.90	8.35
		1.53	1.25	1.02
	Nonart	6.87	8.08	8.08
		1.79	1.03	1.24
Techniques				
	Art	13.18	14.60	15.58
		2.60	1.96	1.75
	Nonart	12.13	14.45	14.70
		2.95	1.82	2.35
Affective intent				
	Art	3.91	4.30	4.40
		1.09	1.11	1.15
	Nonart	3.48	4.13	4.16
		1.48	1.19	1.25
Perceptual sensitivity				
	Art	11.71	13.08	13.38
		2.39	1.98	2.33
	Nonart	10.99	12.74	12.81
		2.90	2.35	2.65
Cognitive deductions				
	Art	3.43	3.87	4.26
		1.12	1.08	.91
	Nonart	3.01	3.77	3.99
		1.26	1.04	1.00

Note. *N*s range from 70 to 347. All values expressed in raw score units. The test component scores were based on a linear combination of the following items: Terms (3, 5, 7, 8, 9, 10, 11, 12, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 31, 37, 38, 40, 41, and 42); Tools (2, 3, 4, 6, 7, 8, 12, 13, and 15); Techniques (3, 4, 6, 7, 8, 12, 13, 14, 15, 20, 24, 25, 38, 40, 41, and 42); Affective intent (25, 33, 34, 35, and 36); Perceptual sensitivity (17, 18, 19, 25, 26, 28, 29, 30, 32, 33, 34, 35, 36, 37, and 38); and Cognitive deduction (4, 6, 20, 23, and 24).

Table 6

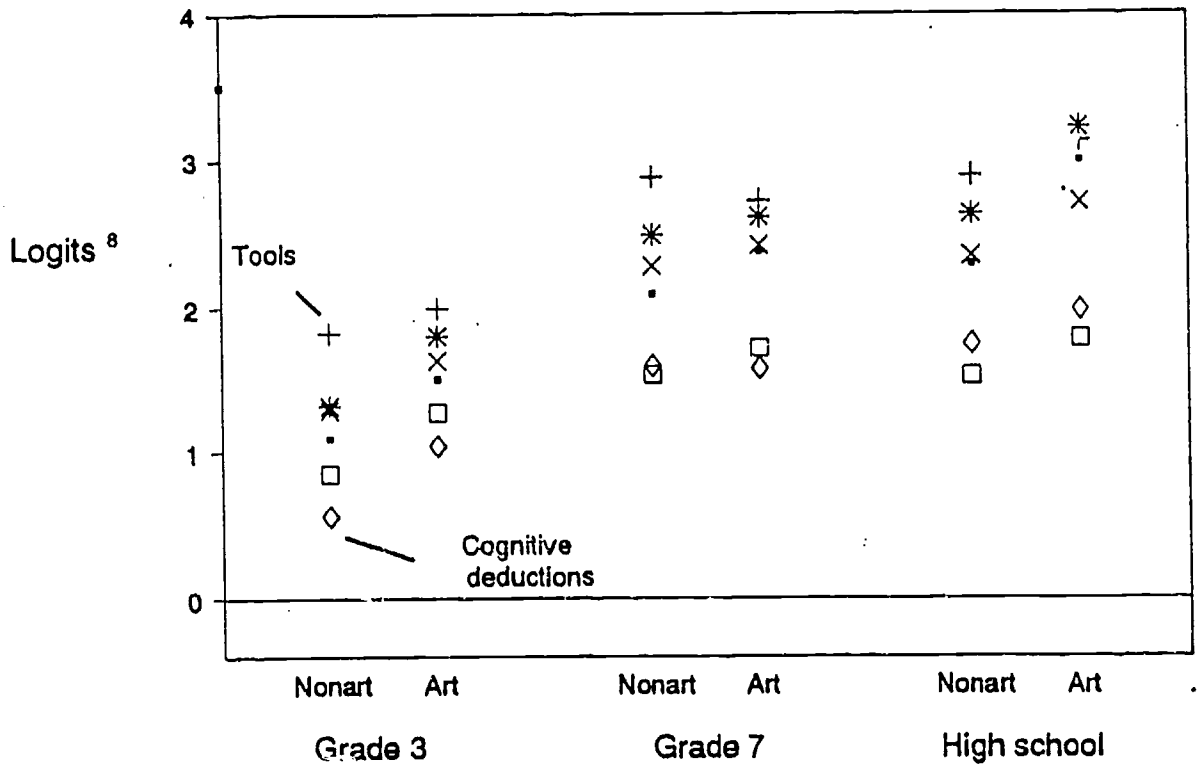
Sum of Squares

Test Component [6]	Sum of Squares	*DF*	Mean Square	*F*	*p*
Terms					
Age	1859.46	2	929.73	94.99	<.001
Education	209.52	1	209.52	21.41	<.001
A X T	34.02	2	17.01	1.74	NS
Error	7419.26	758	9.79		
Tools					
Age	181.77	2	90.89	56.51	<.001
Education	.43	1	.43	.27	NS
A X T	7.02	2	3.51	2.18	NS
Error	1219.13	758	1.61		
Technique					
Age	770.42	2	385.21	80.82	<.001
Education	66.83	1	66.83	14.02	<.001
A X T	24.04	2	12.02	2.52	.08[t]
Error	3612.80	758	4.77		
Affective Intent					
Age	46.94	2	23.47	16.74	<.001
Education	6.33	1	6.33	4.51	<.05
A X T	.62	2	.31	.22	NS
Error	1062.62	758	1.40		
Perceptual Sensitivity					
Age	554.65	2	277.33	45.94	<.001
Education	61.04	1	61.04	10.11	<.001
A X T	5.18	2	2.59	.43	NS
Error	5233.81	758	6.04		
Cognitive Deductions					
Age	80.92	2	40.46	36.03	<.001
Education	12.30	1	12.30	10.96	<.001
A X T	4.14	2	2.21	1.97	NS
Error	851.13	758	1.12		

Note. Total scores were transformed to one-parameter logits. Because many of the items were developmentally inappropriate for the kindergartners (i.e., they were unable to form a valid response), these children were not included in the analysis.

Figure 2

Comparison of Test Component Scores by Art-Educated and Non-Art-Educated Groups [7]



Note. *Ns* range from 79 to 373. The test component scores were based on a linear combination of the following items: Terms (3, 5, 7, 8, 9, 10, 11, 12, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 31, 37, 38, 40, 41, and 42); Tools (2, 3, 4, 6, 7, 8, 12, 13, and 15); Techniques (3, 4, 6, 7, 8, 12, 13, 14, 15, 20, 24, 25, 38, 40, 41, and 42); Affective intent (25, 33, 34, 35, and 36); Perceptual sensitivity (17, 18, 19, 25, 26, 28, 29, 30, 32, 33, 34, 35, 36, 37, and 38); and Cognitive deduction (4, 6, 20, 23, and 24).

however, are not uniform across the test components, and on the component testing knowledge of Tools in grade 7, the non-art group actually scored higher than the art group.

The consistency of the component ordering across grades is somewhat surprising. With the exception of tools, the component ordering between grade 3 and high school does not change, although the difficulty of the knowledge components (i.e., tools, terms, and techniques) tends to become easier relative to the more complex processing components (affective intent, perceptual sensitivity, & cognitive deductions). The results show that as children grow older both art- and non-art-educated learn more about tools and technique and their perceptual processing capabilities improve as well.

Discussion

This study presents a narrow perspective on visual arts assessment by only analyzing the measurement properties, internal structure, and validity of 39 multiple-choice visual arts achievement test items. The results clearly show the items to have good reliability and to be reasonable valid for assessing visual arts learning. These results, however, have implications not only for visual arts assessment, but for visual arts learning in general. They show visual arts instruction to be associated with child responses to test items and visual learning to be characterized by several components of achievement not generally associated with school learning. Consequently, the results suggest unique ways that art education influences child development.

Measurement Properties and Internal Structure

These results support the reliability and validity of a traditional multiple-choice test assessing elementary school visual arts learning. Although specific items may need revision to improve their measurement properties or visual arts validity, a test of these items is remarkably sound. Both items and persons generally fit a linear measurement model, and thus the results show that similar differences in ability represent uniform quantitative differences in achievement. The reliability of the test, for a prototype, is good. The overall sample showed an alpha reliability of .86 and even for the third graders, it was over .80 suggesting that objective evaluations of visual arts learning are possible.

The differences in obtained Rasch fit *t* values (six items showed poor fit in the overall group versus none in grade 3) indicate these items are most effective for third graders, the target population of the assessment but provide useful information about the visual arts learning of older children as well.

Validity

The validity of this approach to assessing learning in the visual arts is supported by several analyses. First, differences between total test scores showed art-educated students to score significantly higher than non-art-educated students, and students with the most art education through elementary school to score the highest. The differences between art- and non-art-educated students first become apparent at the end of kindergarten, but the magnitude of the difference increases with additional years of visual arts education.

A second analysis concerning construct validity, viewed from the difficulty of the internal test components, showed that performance on the items tends to follow a theoretically plausible pattern. Items assessing knowledge of terms and techniques were easier for art trained children; and items assessing perceptual sensitivity, interpretation of an artist's affective intent, and the formation of cognitive deductions become easier as students grow older and acquire additional years of art education. The only component that did not follow this pattern was knowledge of tools which failed to show a significant difference between art-educated and non-art-educated children in grade 7.

Issues in Art Evaluation

The results concerning the use of photographs in test items were encouraging. Although photography always produces some visual distortion of a given image, the high quality color reproductions in this assessment were useful in differentiating the learning of art and non-art-educated children. Even the items that concentrated on physical processes central to artistic production appeared to benefit from visually presented content.

A more important concern is probably the hazard these items present to school art programs. Because this method of assessment forces visual arts achievement into a comparison between what children have learned versus the expectations of a model curriculum, it promotes a segmentation of instruction into learnable objectives that are systematically assessed during evaluation. Consequently, this method of assessment, because of its emphasis on immediately learnable units, may undermine the long range psychological and aesthetic goals of art

appreciation and enjoyment, as well as distort the naturalistic process by which children acquire visual knowledge. The use of this assessment method, in spite of its effectiveness as an evaluation technique, should be undertaken cautiously. It should be integrated into an overall plan for visual arts assessment that may include other sources of student performance, and perhaps emphasized as an instructional tool that indicates mastery of key learning criteria.

The caution expressed above concerning the misuse of this assessment method, however, should not diminish its importance to visual art theory or obscure the opportunities it offers researchers investigating visual arts learning and cognitive development. The results provide substantial empirical evidence that several factors of achievement underlie visual arts learning. This knowledge previously was only the subject of speculation.

Finally, these results sharpen the contrast between the methods of assessment that are now available to art educators (i.e., multiple choice format, performance samples, portfolios, and so on) and increase the importance of understanding the application appropriate for a particular assessment goal.

Dimensions of Ability

Among the most striking results of this study is the empirical delineation of several components of visual arts learning. These components (tools, terms, techniques, affective intent, perceptual sensitivity, and cognitive deductions) show similarities to categories described by other researchers. Machotka (1966), for example, described developmental shifts in the criteria on which children based

their aesthetic judgments, and specifically found that 12 year olds provided more global evaluations of clarity, style, composition, and color than younger children. In other research, Csikszentmihalyi and Robinson (1990) proposed perceptual, emotional, intellectual, and communication characteristics of visual art as major dimensions of aesthetic experience.

The results here provide empirical evidence that categories of art experience are, indeed, important for visual arts achievement and that perceptual, affective, and cognitive components in particular represent important differences between art- and non-art-educated children. The obtained results, however, indicate that the six components of achievement on which this visual arts test was originally based are probably not necessary to describe their responses. A factor analysis found that three primary factors (i.e., perceptual sensitivity, physical sensitivity, and knowledge of tools) were sufficient to describe the responses to the items, and a Rasch measurement analysis showed that these factors can be quantitatively ordered on a continuous variable (knowledge of tools was the easiest, physical and perceptual sensitivity were hardest) with measurement properties of linearity and additivity.

Implications for Art Education

These results raise several issues for art educators. First is a question concerning the appropriateness of the learning objectives in the model visual arts curriculum for grade 3. The difference between art-educated and non-art-educated children in the third grade was modest (< .50 standard deviation units[1]) and not until early adolescence did art-

[1]The mean difference between art and nonart groups was divided by the overall standard deviation.

educated children show a substantial advantage in their test scores. For a variety of reasons (i.e., teachers may not be teaching these objectives, students have difficulty learning them, and so on) some objectives in this curriculum may not be appropriate for third graders. These results suggest that a great deal of visual arts learning probably occurs during elementary schooling as part of children's normal intellectual development and without systematic art education. They encourage art educators to reconsider some of the goals of elementary school visual arts learning.

Summary and Recommendations

1. In general the items testing visual learning met the criterion for linear measurement. The items that showed poor fit tended to test knowledge of visual art terms.

2. Alpha reliability of 39 items for the overall group (N = 777) was .86 and of 38 items for the third graders (N = 250), .81.

3. Total test scores between art and non-art-educated students significantly differed in kindergarten, third grade, seventh grade, and high school. Education and achievement showed a significant interaction in grade 7. As art-educated students increased their art background, their test scores increased. Art-educated students showed a significantly higher achievement in grade 7.

4. An analysis of the test component scores shows that the component assessing knowledge of tools had many of the easiest items

and the component assessing Cognitive Deductions had the most difficult items.

5. An analysis of test component scores for art-educated and non-art-educated students showed that students in grade 7 do not significantly differ in their knowledge of tools.

References

Arnheim, R. (1969). *Visual thinking.* Berkeley: University of California Press.

Arnheim, R. (1986). *New Essays on the Psychology of Art.* Berkeley: University of California Press.

Bennett, G. (1989). Words in my eyes: The assessment of art and design in GCSE. *Educational Studies, 15,* 241-249.

Bezruczko, N. (1989). *An Assessment of Visual Learning.* Chicago: Sheffield Press.

Chicago Board of Education (1991). *The state report cards for 1990-1991: An analysis of the first six years.* Chicago: The Board.

Chicago Board of Education (1989). *1988-1989 Test scores and selected school characteristics: Elementary Schools.* Chicago: The Board.

Csikszentmihalyi, M., & Robinson, R. (1990). *The art of seeing: An interpretation of the aesthetic encounter.* Malibu, CA: The Getty Center for Education in the Arts.

Ecker, D.W. (1963). The artistic process as qualitative problem solving. *Journal of Aesthetics and Art Criticism, 21,* 283-290.

Csikszentmihalyi, M., & Getzels, J. (1988). Creativity and problem finding in art. In F. H. Farley & R. W. Neperud (Eds.), *The foundations of aesthetics, art, and art education* (pp. 91-116). New York: Praeger.

Eisner, E.W. (1985). *The art of educational evaluation: A personal view.* London: The Falmer Press.

Frechtling, J.A. (1991). Performance assessment: Moonstruck or the real thing? *Educational measurement, 10,* 23-25.

Gardner, H. (1982). *Art, mind, and brain.* New York: Basic Books, inc.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books, inc.

Getty Institute for Educators on the Visual Arts. (1985). *Executive summary of the final report on Institute '84.* Los Angeles: Author.

Getzels, J., & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art.* New York: John Wiley & Sons.

Hamblen, K.A. (1992). *Research and theories supporting art instruction for instrumental outcomes.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Higgins, R.E. (1989). *Approaches to evaluation in art education at the state level.* ERIC Document No. ED310157.

Hoepfner, R. (1984). Measuring student achievement in art. *Studies in art education, 25,* 251-258.

Illinois State Board of Education. (n.d.) *State Goals for Learning and Sample Learning Objectives: Fine Arts; Grades 3, 6, 8, 10, 12.* Springfield, Illinois: Illinois State Board of Education Department of School Improvement Services.

Machotka, P. (1966). Aesthetic criteria in childhood: Justifications of preference. *Child Development, 37,* 876-885.

Mitchell, J.V. (1985). (Ed.) *The ninth mental measurements yearbook.* Lincoln, Nebraska: University of Nebraska Press, Buros Institute of Mental Measurements.

National Assessment of Educational Progress. (1981). *Art and young Americans, 1974-1979: Results from the second national art assessment.* Art Report No. 10-A-01. Washington: Author.

National Assessment of Educational Progress. (1978a). *Attitudes toward art: Selected results from the first national assessment of art.* Art Report No. 06-A-03. Washington: Author.

National Assessment of Educational Progress. (1978b). *Knowledge about art: Selected results from the first national assessment of art.* Art Report 06-A-02. Washington: Author.

Wright, B.D., & Linacre, J.M. (1992). *Bigsteps: A Rasch program for rating scale analysis.* Chicago: MESA Press.

Wright, B.D., & Stone, M. (1979). *Best test design.* Chicago: MESA Press.

Appendix

Curriculum Objectives and Internal Structure

Item No.	Curriculum [9] Objective	P-value	Item-total correlations	Rasch [10] logit	Fit t [11] value
2	Identify or use materials as simple looms	.68	.27	1.08	3.3
3	Identify or use weaving and stitchery materials	.90	.34	-.77	-.8
4	Identify or use common objects for print making	.87	.27	-.37	1.7
5	Identify primary and secondary colors	.58	.29	1.69	4.3
6	Identify simple movement and direction of line	.57	.37	1.71	1.0
7	Identify or use basic drawing and painting tools	.95	.27	-1.64	-.3
8	Identify or use basic drawing and painting tools	.97	.23	-2.31	-.6
9	Know or demonstrate how to work with a variety of tools	.97	.27	-2.18	-1.0
10	Know or demonstrate how to work with a variety of tools	.75	.26	.68	3.6
11	Identify or use sculpting materials	.97	.20	-2.17	-.8
12	Know or demonstrate how to work with a variety of tools	.91	.32	-.78	-.4
13	Identify or use weaving and stitchery materials	.89	.36	-.65	-1.5
14	Identify found objects which can be used for jewelry	.96	.20	-1.88	-.3
15	Know or demonstrate how to work with a variety of tools	.89	.40	-.58	-1.4
17	Recognize the qualities of texture	.92	.27	-.96	-.4
18	Identify simple movement and direction of line	.96	.32	-1.88	-.7
19	Understand how choices of colors combine with other elements in that image	.69	.35	1.03	1.3

Curriculum Objectives and Internal Structure (continued)

Item No.	Curriculum Objective	P-value	item-total	Rasch logit	Fit *t* value
20	Understand how choices of technique and tools combine with other elements in that image	.55	.32	1.83	3.5
21	Identify geometric, irregular, and natural shapes	.87	.32	-.36	.0
22	Identify geometric, irregular, and natural shapes	.39	.18	2.73	5.5
23	Discriminate between figure and ground	.91	.35	-.90	-.4
24	Understand how choices of materials and media combine with other elements in that image	.76	.42	.62	-2.0
25	Understand how media choice conveys atmosphere	.85	.40	-.16	-1.1
26	Identify or use sculpting materials	.72	.34	.87	.6
27	Identify given significant visual images	.57	.33	1.76	3.1
28	Identify similarities/differences in size, shape, color, and tone	.91	.32	-.87	-1.0
29	Recognize the qualities of texture	.84	.38	-.05	-.9
30	Recognize expressive and physical characteristics of color	.90	.24	-.77	.0
31	Identify formal, informal, and radial balance	.83	.23	.07	1.2
32	Identify rhythm	.89	.48	-.55	-2.2
33	Relate the mood	.74	.47	.74	-2.3
34	Understand the emotions communicated through facial expressions and actions	.78	.37	.41	-.2
35	Understand the emotions communicated through facial expressions and actions	.74	.46	.76	-2.0

Curriculum Objectives and Internal Structure (continued)

Item No.	Curriculum Objective	P-value	item-total	Rasch logit	Fit *t* value
36	Relate the mood	.87	.33	-.38	-.4
37	Identify simple movement and direction of line	.62	.45	1.45	-2.4
38	Understand ways the sensory, formal, and technical qualities perceived in an art work interact to express ideas	.74	.33	.77	1.0
40	Identify or demonstrate simple printing processes	.45	.35	2.35	1.3
41	Identify or demonstrate paper construction processes such as curling, slotting, or folding	.84	.41	.07	-.7
42	Identify or demonstrate or use basic drawing and painting tools	.88	.41	-.40	-1.0

Note: *N*s for the items range from 912 to 993. Students in grades K, 3, 7, and high school are included in the sample.

Author Note

The instrument that was evaluated in this report was prepared by Nikolaus Bezruczko for the Board of Education of the City of Chicago under contract with the Illinois State Board of Education. Further information concerning this test of visual arts achievement can be obtained from the author at 1532 E. 59th Street, Chicago, Illinois 60637, U.S.A.

Thanks are due to the Chicago Public Schools, Bureau of Student Testing for supporting the development of an objective method of assessing visual arts achievement. My special thanks to Carole L. Perlman, Director, Bureau of Student Testing for her participation in several reviews of the items and her general cooperation throughout the project.

The cooperation and resources of the teachers and principals who participated in this project are gratefully acknowledged. Lucinda Vriner, an art teacher at the Franklin Fine Arts Academy, was an extraordinary resource, as well as collaborator, throughout the undertaking and especially during the development of items and their field tryout. Her expertise and enthusiasm during the construction of the items helped mediate the discouragement of the early item trials.

I am especially grateful to Susan Friefeld of the Terra Museum of American Art for her interest in visual arts assessment. The items in this test could not have been physically produced without her active participation.

Financial support for the development of this instrument was provided through a grant by the Illinois State Board of Education Department of Program Development and Delivery. Financial support for

a study of reliability, validity, and the preparation of this report was provided by the author.

The interpretation of the results in this report and the recommendations presented do not necessarily represent those of the Illinois State Board of Education Department of Program Development and Delivery or the teachers and administrators of the Chicago Public Schools.

Portions of this study were presented at the 1992 Annual Meeting of the American Educational Research Association, San Francisco.

End Notes

1. Describes the neighborhood immediately contiguous to a school: P (urban poor and nonwhite), A(urban affluent and white), W (urban working class and non-Anglo-white).

2. One-parameter logistic scale values were estimated using Bigsteps (Wright & Linacre, 1992).

3. An unweighted infit statistic (Wright & Stone, 1979) was used to assess fit of items and persons to the Rasch measurement model.

4. Total test scores were transformed to one-parameter logit scale values.

5. Because many of the items were developmentally inappropriate for the kindergartners (i.e., they were unable to form a valid response), these children were not included in this comparison.

6. The test component scores were based on a linear combination of the following items: Terms (3, 5, 7, 8, 9, 10, 11, 12, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 31, 37, 38, 40, 41, and 42); Tools (2, 3, 4, 6, 7, 8, 12, 13, and 15); Techniques (3, 4, 6, 7, 8, 12, 13, 14, 15, 20, 24, 25, 38, 40, 41, and 42); Affective intent (25, 33, 34, 35, and 36); Perceptual sensitivity (17, 18, 19, 25, 26, 28, 29, 30, 32, 33, 34, 35, 36, 37, and 38); and Cognitive deduction (4, 6, 20, 23, and 24).

7. Because many of the items were developmentally inappropriate for the kindergartners (i.e., they were unable to form a valid response), these children were not included in this comparison.

8. Total component scores transformed to one-parameter logits estimated on the art-trained and non-art-trained groups separately.

9. The sample learning objectives appear in *State Goals for Learning and Sample Learning Objectives: Fine Arts; Grades 3, 6, 8, 10, 12* published by the Illinois State Board of Education Department of School Improvement Services.

10. One-parameter logistic scale values were estimated using Bigscale (Wright, 1989).

11. An unweighted infit statistic (Wright & Stone, 1979) was used to assess fit of items and persons to the Rasch measurement model.