



ETS NAEP TECHNICAL AND RESEARCH REPORT SERIES

The NAEP 1997 Arts Technical Analysis Report

Nancy L. Allen
Frank Jenkins
Terry L. Schoeps

in collaboration with

Nancy W. Caldwell, Lucy M. Gray, Laura J. Jerry, Edward Kulick,
Debra M. Kline, Venus Leung, Alfred M. Rogers, Connie Smith,
Spencer S. Swinton, and Xiaohui Wang

March 2004
Technical Report
ETS-NAEP 04-T01

*Listening.
Learning.
Leading.*

The NAEP 1997 Arts Technical Analysis Report

Nancy L. Allen
Frank Jenkins
Terry L. Schoeps

in collaboration with

Nancy W. Caldwell, Lucy M. Gray, Laura J. Jerry, Edward Kulick,
Debra M. Kline, Venus Leung, Alfred M. Rogers, Connie Smith,
Spencer S. Swinton, and Xiaohui Wang

March 2004

Technical Report

ETS-NAEP 04-T01

ETS-NAEP Research and Technical Reports provide limited dissemination of
ETS research on National Assessment of Educational Progress topics.

They are available without charge from:

Research Publications Office

Mail Stop 7-R

ETS

Princeton, NJ 08541

As well as on the worldwide web at:

<http://www.ets.org/research/allreports.html>

The work reported herein was supported under the Contract Award No. ED-02-CO-0023 from the National Center for Education Statistics (NCES), within the Institute of Education Sciences (IES) of the U.S. Department of Education. The NCES project officer is Arnold Goldstein.

- TABLE OF CONTENTS -

Introduction	<i>Nancy L. Allen & James E. Carlson, Educational Testing Service</i>	1
Section 1	Overview of the NAEP 1997 Arts Assessment <i>Alfred M. Rogers, Fred Yan, Terry L. Schoeps, and Debra M. Kline, Educational Testing Service</i>	
1.1	School and Student Samples	3
1.2	Assessment Design.....	3
1.3	Assessment Questionnaires	5
1.4	Data Collection	5
1.5	Scoring	6
1.6	Reporting NAEP Arts Results.....	6
1.7	Organization of the <i>NAEP 1997 Arts Technical Analysis Report</i>	7
Section 2	Special Considerations for the NAEP 1997 Arts Assessment Analyses <i>Alfred M. Rogers, Fred Yan, Terry L. Schoeps, and Debra M. Kline, Educational Testing Service</i>	
2.1	Student and School Samples	9
2.2	Theatre Teacher Response Data	12
2.3	School Response Data	12
2.4	SD/LEP Data	12
2.5	Using Weights in Analyses of NAEP Arts Data	13
2.6	Student Assessment Instruments	15
2.7	NAEP Reporting Groups	19
2.8	Derived Variables	22
2.9	NAEP Arts Scales and Scores.....	23
2.10	Drawing Inferences from the Results.....	25
Section 3	Data Analysis for the 1997 Arts Assessment in Music <i>Frank Jenkins, Spencer S. Swinton, Laura J. Jerry, Edward Kulick, and Shuyi Hua</i> <i>Educational Testing Service</i>	
3.1	Introduction.....	29
3.2	Creating the Item Response Theory (IRT) Scale for Music <i>Responding</i> Items.....	29
3.3	Calculating the Mean Percent-Correct Scales for Music <i>Creating/Performing</i> Items.....	37
Section 4	Data Analysis for the 1997 Arts Assessment in Theatre <i>Frank Jenkins, Spencer S. Swinton, Laura J. Jerry, Edward Kulick, and Venus Leung</i> <i>Educational Testing Service</i>	
4.1	Introduction.....	41
4.2	Creating the Item Response Theory (IRT) Scale for Theatre <i>Responding</i> Items.....	41
4.3	Calculating the Mean Percent-Correct Scale for Theatre <i>Creating/Performing</i> Items.....	47
Section 5	Data Analysis for the 1997 Arts Assessment in Visual Arts <i>Frank Jenkins, Spencer S. Swinton, Laura J. Jerry, Edward Kulick, and Xiaohui Wang</i> <i>Educational Testing Service</i>	
5.1	Introduction.....	51
5.2	Creating the Item Response Theory (IRT) Scale for Visual Arts <i>Responding</i> Items	51
5.3	Calculating the Mean Percent-Correct Scale for Visual Arts <i>Creating</i> Items	57

Appendix A	NAEP 1997 Arts Assessment Sample Design and Weighting Procedures	
A.1	Introduction.....	61
A.2	Primary Sampling Units.....	61
A.3	Selection of Schools.....	62
A.4	Assignment of Sessions to Schools.....	63
A.5	Sampling Students.....	63
A.6	Excluded Students.....	65
A.7	School and Student Participation Rates.....	65
A.8	Overall Student Participation Rates.....	66
A.9	Teacher Questionnaires.....	67
Appendix B	NAEP 1997 Arts Assessment Field Operations and Data Collection	
B.1	Introduction.....	73
B.2	Organization of the Arts Assessment.....	73
B.3	Preparing for the Assessments.....	75
B.4	Selecting the Student Samples.....	79
B.5	Conducting the Assessment Sessions.....	80
B.6	Results of the NAEP Arts Assessment.....	82
B.7	Field Management.....	84
Appendix C	NAEP 1997 Arts Assessment Processing and Scoring Assessment Materials	
C.1	Introduction.....	85
C.2	Overview.....	85
C.3	Packing and Distribution.....	86
C.4	Processing of Test Materials.....	89
C.5	Scoring Overview.....	98
C.6	Preparation for Tape Creation.....	105
C.7	Uploading of Scores to Database.....	105
C.8	SD/LEP Questionnaires.....	106
C.9	School Questionnaires.....	106
C.10	Teacher Questionnaire Match.....	106
C.11	Delivery & Storage of Documents.....	106
C.12	Quality-Control Documents.....	107
Appendix D	NAEP 1997 Arts Assessment Scoring Reliability	109
Appendix E	NAEP 1997 Arts Assessment Items Contributing to Each Scale	121
Appendix F	NAEP 1997 Arts Assessment IRT Parameters	125
Appendix G	NAEP 1997 Arts Assessment Summary Tables of Variables Used to Define Groups of Students	131
Reference List	151

LIST OF TABLES

Introduction	(No tables)	
Section 1	Overview of the NAEP 1997 Arts Assessment	(No tables)
Section 2	Special Considerations for the NAEP 1997 Arts Assessment Analyses	
	Table 2-1 NAEP 1997 Arts Assessment School, Student, and Teacher Questionnaire Participation Rates.....	11
	Table 2-2 NAEP 1997 Arts Assessment Number of Assessed Students by Sample and Item Type	15
	Table 2-3 NAEP 1997 Arts Assessment Total Number of Blocks by Arts Discipline	15
	Table 2-4 NAEP 1997 Arts Assessment Distribution of Questions by Item Type	16
	Table 2-5 NAEP 1997 Arts Book Map – Music	17
	Table 2-6 NAEP 1997 Arts Book Map – Theatre.....	17
	Table 2-7 NAEP 1997 Arts Book Map – Visual Arts.....	18
	Table 2-8 NAEP 1997 Arts Score Scales.....	25
Section 3	Data Analysis for the 1997 Arts Assessment in Music	
	Table 3-1 NAEP 1997 Arts Assessment Descriptive Statistics for the IRT-Scaled <i>Responding</i> Items by Block for the Music Sample.....	30
	Table 3-2 NAEP 1997 Arts Assessment Score Levels for Polytomous Music Items	32
	Table 3-3 NAEP 1997 Arts Assessment Music Items Receiving Special Treatment	32
	Table 3-4 NAEP 1997 Arts Assessment in Music – Proportion of Scale Score Variance Accounted For by the Estimation Model.....	36
	Table 3-5 NAEP 1997 Arts Assessment Means and Standard Deviations of All Five Plausible Values for the Music <i>Responding</i> Scales	36
	Table 3-6 NAEP 1997 Arts Assessment Transformation Constants for the Music <i>Responding</i> Scales.....	37
	Table 3-7 NAEP 1997 Arts Assessment Descriptive Statistics for the <i>Creating</i> and <i>Performing</i> Items by Block for the Music Sample	38
	Table 3-8 NAEP 1997 Arts Assessment Session/Block Structure for Music	39
	Table 3-9 NAEP 1997 Arts Assessment Items in the Music <i>Creating</i> Scale.....	40
	Table 3-10 NAEP 1997 Arts Assessment Items in the Music <i>Performing</i> Scale.....	40
Section 4	Data Analysis for the 1997 Arts Assessment in Theatre	
	Table 4-1 NAEP 1997 Arts Assessment Descriptive Statistics for the IRT-Scaled <i>Responding</i> Items by Block for the Theatre Sample	42
	Table 4-2 NAEP 1997 Arts Assessment Score Levels for Polytomous Theatre Items	44
	Table 4-3 NAEP 1997 Arts Assessment Theatre Items Receiving Special Treatment	45
	Table 4-4 NAEP 1997 Arts Assessment in Theatre Proportion of Scale Score Variance Accounted for by the Estimation Model.....	46
	Table 4-5 NAEP 1997 Arts Assessment Means and Standard Deviations of All Five Plausible Values for the Theatre <i>Responding</i> Scale	46
	Table 4-6 NAEP 1997 Arts Assessment Transformation Constants for the Theatre <i>Responding</i> Scale	47
	Table 4-7 NAEP 1997 Arts Assessment Descriptive Statistics for the <i>Creating/Performing</i> Items by Block for the Theatre Sample.....	48
	Table 4-8 NAEP 1997 Arts Assessment Session/Block Structure for Theatre	48
	Table 4-9 NAEP 1997 Arts Assessment Items in the Theatre <i>Creating/Performing</i> Scale	49

Section 5	Data Analysis for the 1997 Arts Assessment in Visual Arts	
	Table 5-1 NAEP 1997 Arts Assessment Descriptive Statistics for the IRT-Scaled <i>Responding</i> Items by Block for the Visual Arts Sample	52
	Table 5-2 NAEP 1997 Arts Assessment Score Levels for Polytomous Visual Arts Items	54
	Table 5-3 NAEP 1997 Arts Assessment Visual Arts Items Receiving Special Treatment	54
	Table 5-4 NAEP 1997 Arts Assessment in Visual Arts Proportion of Scale Score Variance Accounted for by the Estimation Model	56
	Table 5-5 NAEP 1997 Arts Assessment Means and Standard Deviations of All Five Plausible Values for the Visual Arts <i>Responding</i> Scale	56
	Table 5-6 NAEP 1997 Arts Assessment Transformation Constants for the Visual Arts <i>Responding</i> Scale	57
	Table 5-7 NAEP 1997 Arts Assessment Descriptive Statistics for the Visual Arts <i>Creating</i> Items by Block	57
	Table 5-8 NAEP 1997 Arts Assessment Session/Block Structure for Visual Arts	59
	Table 5-9 NAEP 1997 Arts Assessment Items Contributing to the Visual Arts <i>Creating</i> Scale	60
Appendix A	NAEP 1997 Arts Assessment Sample Design and Weighting Procedures	
	Table A-1 NAEP 1997 Arts Assessment Weighted Participation Rates of Sampled Schools by Public/Private Status	65
	Table A-2 NAEP 1997 Arts Assessment Weighted Participation Rates of Sampled Students After Removing Excluded Students.....	66
	Table A-3 NAEP 1997 Arts Assessment Overall Student Participation Rates	66
	Table A-4 NAEP 1997 Arts Assessment Post-Stratification Totals	71
Appendix B	NAEP 1997 Arts Assessment Field Operations and Data Collection	
	Table B-1 NAEP 1997 Assessment Schedule of Field Activities.....	75
	Table B-2 NAEP 1997 Assessment Questionnaires Distributed and Completed	83
Appendix C	NAEP 1997 Arts Assessment IRT Parameters	
	Table C-1 NAEP 1997 Arts Assessment Participation Counts.....	89
	Table C-2 NAEP 1997 Grade 8 Arts Assessment and Grade 12 Arts Field Test Dates	100
	Table C-3 NAEP 1997 Grade 8 Arts Assessment Number of Constructed-Response Items.....	102
	Table C-4 NAEP 1997 Grade 8 Arts Assessment Inter-Reader Reliability Ranges.....	104
Appendix D	NAEP 1997 Arts Assessment Scoring Reliability	
	Table D-1 NAEP 1997 Arts Assessment Range of Response Codes, Percent Agreement, and Cohen's Kappa for the Dichotomously Scored Constructed-Response Music Items	110
	Table D-2 NAEP 1997 Arts Assessment Range of Response Codes, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Music Items	111
	Table D-3 NAEP 1997 Arts Assessment Range of Response Codes, Percent Agreement, and Cohen's Kappa for the Dichotomously Scored Constructed-Response Theatre Items	114
	Table D-4 NAEP 1997 Arts Assessment Range of Response Codes, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Theatre Items	115
	Table D-5 NAEP 1997 Arts Assessment Range of Response Codes, Percent Agreement, and Cohen's Kappa for the Dichotomously Scored Constructed-Response Visual Arts Items.....	117

Table D-6	NAEP 1997 Arts Assessment Range of Response Codes, Percent Agreement, and Intraclass Correlation for the Polytomously Scored Constructed-Response Visual Arts Items.....	118
-----------	--	-----

Appendix E NAEP 1997 Arts Assessment Items Contributing to Each Arts Scale

Table E-1	NAEP 1997 Arts Assessment Items Contributing to the Music <i>Creating</i> Scale.....	121
Table E-2	NAEP 1997 Arts Assessment Items Contributing to the Music <i>Performing</i> Scale.....	121
Table E-3	NAEP 1997 Arts Assessment Items Contributing to Music <i>Responding</i> Scale 1.....	122
Table E-4	NAEP 1997 Arts Assessment Items Contributing to Music <i>Responding</i> Scale 2.....	122
Table E-5	NAEP 1997 Arts Assessment Items Contributing to the Theatre <i>Creating/Performing</i> Scale.....	123
Table E-6	NAEP 1997 Arts Assessment Items Contributing to the Theatre <i>Responding</i> Scale.....	123
Table E-7	NAEP 1997 Arts Assessment Items Contributing to the Visual Arts <i>Creating</i> Scale.....	124
Table E-8	NAEP 1997 Arts Assessment Items Contributing to the Visual Arts <i>Responding</i> Scale.....	124

Appendix F NAEP 1997 Arts Assessment IRT Parameters

Table F-1	NAEP 1997 Arts Assessment IRT Parameters for Music Scale 1 Sample.....	126
Table F-2	NAEP 1997 Arts Assessment IRT Parameters for Music Scale 2 Sample.....	127
Table F-3	NAEP 1997 Arts Assessment IRT Parameters for the Theatre Sample.....	128
Table F-4	NAEP 1997 Arts Assessment IRT Parameters for the Visual Arts Sample.....	129

Appendix G NAEP 1997 Arts Assessment Summary Tables of Variables Used to Define Groups of Students

Table G-1	NAEP 1997 Arts Assessment Description of Specifications Provided for Each Variable Defining Group Membership.....	132
Table G-2	Summary Table of the 1997 Music Variables Defining Group Membership.....	132
Table G-3	Summary Table of the 1997 Theatre Variables Defining Membership.....	137
Table G-4	Summary Table of the 1997 Visual Arts Variables Defining Group Membership.....	145

Reference List (No tables)

INTRODUCTION

*Nancy L. Allen and James E. Carlson
Educational Testing Service*

As the nation's only ongoing survey of students' educational progress, the National Assessment of Educational Progress (NAEP) is an important resource for understanding what students know and can do. NAEP assessments have explored students' abilities in a range of subject areas, including reading, writing, mathematics, science, U.S. history, and world geography. Based on assessment results, NAEP reports levels of student achievement and instructional, institutional, and demographic variables associated with those levels of achievement.

The 1997 National Assessment of Educational Progress monitored the performance of students in the arts disciplines of Music, Theatre, and Visual Arts. The national main samples involved public and nonpublic school students who were in the eighth grade. Nearly 6,700 students from 268 schools were assessed.

The purpose of this report is to provide details on the data analysis procedures for the 1997 assessment. It also contains details on the instruments, sample design, and data collection that influenced analysis decisions. This report provides information necessary to show adherence to the National Center for Education Statistic's *NCES Statistical Standards* (NCES, 2003); the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1985) and to the Educational Testing Service (ETS) *Standards for Quality and Fairness* (Educational Testing Service, 2003).

Detailed substantive results are not presented here but can be found in a series of NAEP reports covering the status of student performance; several additional technical memos or reports provide information on how the assessment was designed and implemented. For each of the arts disciplines, *The NAEP 1997 Arts Report Card* (Persky, Sandene, & Askew, 1998) describes the achievement of eighth graders within the general population and in various subgroups. Taken with the information provided about instructional and institutional variables, this provides a context for evaluating the status of students' learning in the arts. In addition to the printed *Report Card*, a CD-ROM is available that features the complete text of the report, as well as many examples of assessment exercises and student responses. Additional information is provided in the *1997 Arts Education Assessment Framework* (National Assessment Governing Board [NAGB], 1994); *Focus on NAEP: The NAEP 1997 Arts Education Assessment: An Overview* (White & Vanneman, 1998a); and a series of four *Focus on NAEP* publications, short descriptions about the content and process of each of the four different 1997 Arts Assessments: dance, music, theatre, and visual arts (Vanneman & Goodwin, 1998; Vanneman, Morton, & Allen, 1998; Vanneman, Schuler, & Sandene, 1998; White & Vanneman, 1998b). The *NAEP 1997 Arts Data Companion* (Rogers, Yan, Kline, & Schoeps, 2000) provides information required for analysis of the 1997 NAEP results, and the *NAEP Guide: A Description of the Content and Methods of the 1997 and*

1998 Assessments (Calderone, King, & Horkay, 1997) describes the content and methods used in the 1997 assessment.

Many of the NAEP reports, including summary data tables, are available on the Internet at <http://nces.ed.gov/nationsreportcard>. For information about ordering printed copies of these reports, go to the Department of Education Web Page at <http://www.ed.gov/pubs/edpubs.html>, call toll free 1-877-4ED PUBLS (877-433-7827), or write to:

Education Publications Center (ED Pubs)
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794 –1398

For ordering information on the assessment frameworks, write to:

National Assessment Governing Board
800 North Capitol Street NW
Suite 825
Washington, DC 20002

The frameworks and other NAGB documents are also available through the Internet at <http://www.nagb.org>.

Section 1

OVERVIEW OF THE NAEP 1997 ARTS ASSESSMENT¹

*Alfred M. Rogers, Fred Yan, Terry L. Schoeps, and Debra M. Kline
Educational Testing Service*

In 1997, NAEP conducted a national assessment at grade 8 in the arts disciplines of Music, Theatre, and Visual Arts. (Though an assessment was developed for dance, it was not implemented because a statistically suitable sample could not be located.) The assessment was conducted by Educational Testing Service (ETS) for the National Center for Education Statistics (NCES) and funded by the U.S. Department of Education. ETS was responsible for overall management of the program, development of the design, development of the items and questionnaires, data analysis, and reporting. Westat was responsible for all aspects of sampling and field operations, while Pearson Educational Measurement carried out the printing, distribution, and receipt of materials; the scanning of assessment data; and the professional scoring of constructed responses.

1.1 School and Student Samples

The NAEP 1997 Arts Assessment was conducted nationally at grade 8. For Music and Visual Arts, representative samples of public and nonpublic school students were assessed. The Arts Assessment in Theatre was administered to a “targeted” sample of students selected only from schools with theatre programs. For Music, 2,275 students were assessed (1,999 from public schools and 276 from nonpublic schools). For Theatre, 1,386 students were assessed (1,335 from public schools and 51 from nonpublic schools); however, the sample of nonpublic schools for Theatre was not large enough to permit the separate reporting of nonpublic school results. For Visual Arts, 2,999 students were assessed (2,756 from public schools and 243 from nonpublic schools). Section 2 contains information on sample sizes and participation rates for the assessment.

For the Arts Assessment in Music and Visual Arts, the national and regional data are based on nationally representative probability samples of all eighth-grade students. The results for Theatre are based on a representative sample of eighth-grade theatre students representing those schools in the nation with an extensive theatre curriculum. The samples were selected using a complex multistage sampling design that involved sampling students from selected schools within selected geographic areas across the country (see Section 2 and Appendix A for more information).

1.2 Assessment Design

The *1997 Arts Education Assessment Framework* set by the National Assessment Governing Board (NAGB, 1994) was built around the three arts processes of *Responding*, *Creating*, and *Performing*.

¹ Nancy L. Allen edited Section 1 for use in this report.

In Music, the processes of *Responding*, *Creating*, and *Performing* are all emphasized, although music educators have typically placed greater emphasis on performance of existing works and students' responses to performances. In Theatre, *Creating* and *Performing* are understood as a combined act, and the responses of the audience, director, actors, and designers are seen as important components of the development of *Creating/Performing* work. In Visual Arts, creative expression and responses to artworks are more highly valued than the performance, or duplication, of existing works.

Independent sets of exercises were developed for each discipline of the three assessed arts disciplines. Within each discipline, there were four "blocks" (groups of exercises administered as separate units to be completed in a set time frame) of written *Responding* exercises, and three *Creating* and/or *Performing* blocks. (In Theatre, *Creating* and *Performing* constitute one category, and Visual Arts do not include the process of *Performing*.) The Arts Assessment in Music also featured two additional *Creating* and *Performing* blocks designed for students who indicated that they were currently involved in some musical activity.

Each student who participated in the assessment was assessed in one of the three arts disciplines, to ensure that sufficiently in-depth information about students' arts abilities was gathered. In the first portion of the assessment, each student received one booklet containing two blocks of cognitive *Responding* (also called A/B) exercises and two blocks of background questions. The cognitive blocks included multiple-choice questions and two types of constructed-response questions: short constructed-response questions that required students to write answers of a few words or sentences, and extended constructed-response questions that required students to provide answers of a paragraph or more. (In Visual Arts, three of the *Responding* blocks (also called A/B blocks) included two-dimensional *Creating* tasks.) Answers to the constructed-response questions were evaluated using multi-level scoring guides that defined criteria for full credit, partial credit, or no credit. The background questions asked students to provide information about their demographic characteristics, arts classroom instruction, and self-perceptions about their abilities in the art form in which they were being assessed.

In the second portion of the assessment, each student completed one *Creating* and/or *Performing* block. (The exception is Music. Students who were currently engaged in some type of musical activity took two *Creating* and/or *Performing* blocks, one for the general student sample and one for students with special music knowledge.) No background questions were asked during this part of the assessment.

Separating the more active portions of the assessment allowed a suitable amount of time to be devoted to these complex tasks. It also allowed for numerous special conditions that had to be met to successfully administer *Creating* and/or *Performing* tasks. Among these were the need to have students work in pairs or groups for Theatre improvisations; the need to videotape students acting; the need to set up instruments and recording devices for Music tasks; the complications associated with distributing large amounts of Visual Arts materials; the time needed to photograph three-dimensional Visual Arts works for future scoring; and in general, special space requirements for all three arts.

Section 2 provides more detail about the student assessment instruments.

1.3 Assessment Questionnaires

In addition to the assessment exercises administered to each student, each booklet in the assessment also included several sets of background questionnaires. Students sampled for the Arts Assessment completed one 5-minute set of student demographic background questions and one 10-minute set of subject-specific background questions. The subject-specific background questionnaires were designed to gather contextual information about students, their instructional and out-of-school arts experiences, and their attitudes toward the art domain in which they were being assessed.

To supplement the information on instruction reported by students, the theatre teachers of the targeted students participating in the NAEP Arts Assessment in Theatre were asked to complete a questionnaire about instructional practices, teaching backgrounds, and characteristics. The results of the field tests in Visual Arts and Music showed high percentages of missing data for the Visual Arts and Music teachers' questionnaires. Because of this, teacher questionnaires were not administered in the operational Arts Assessments in Visual Arts and Music.

The principals of students sampled for the assessment were asked to complete a questionnaire about the school's characteristics and students' access to instruction in the arts.

An additional questionnaire was designed to gather information about students with disabilities (SD) and limited-English proficient (LEP) students.

A more extensive discussion of the student, (theatre) teacher, school, and SD/LEP questionnaires is provided in Section 2.

1.4 Data Collection

In addition to sample selection, Westat was responsible for field administration and data collection for the 1997 Arts Assessment. When data collection was completed, assessment instruments were sent to National Computer Systems for processing and professional scoring. The resulting data files were then sent to Educational Testing Service (ETS), where they were transcribed to a database ready for analysis. Additional information on data collection is provided in Appendix B.

From mid-September to mid-December 1996, NAEP/Westat field staff contacted districts for cooperation, and conducted introductory meetings, if needed. Exercise administrators were hired and trained by the assessment supervisors in mid-March 1997 immediately prior to the data collection period. The grade 8 Arts Assessments were administered between March 24 and May 9, 1997.

Under the direction of the Westat home office staff, Westat field managers led the activities of assessment supervisors, who oversaw the work of the exercise administrators as they visited each school.

The assessments in each participating school were coordinated by the assessment supervisors, who were responsible for conducting sessions in their primary sampling units² (PSUs) and maintaining the security of NAEP materials and the confidentiality of assessment data. Assessment supervisors

² NAEP primary sampling units are metropolitan statistical areas, counties, or groups of contiguous counties in the United States.

provided school personnel with general information about NAEP and worked with them in an effort to maximize student attendance at each assessment session. Supervisors and exercise administrators were responsible for establishing assessment schedules, ensuring that schools were prepared for sessions, completing student sample selection in each school, collecting and checking all NAEP materials, and returning the completed assessment materials to Pearson Educational Measurement for processing.

1.5 Scoring

Materials from the NAEP 1997 Arts Assessment were shipped to Pearson Educational Measurement, where trained staff evaluated the responses to the constructed-response questions using scoring rubrics prepared by Educational Testing Service. Each constructed-response question had a unique scoring guide that defined the criteria used to evaluate students' responses. The extended constructed-response questions generally were evaluated with four- or five-level scoring guides, while the short constructed-response questions generally were scored with two- and three-level scoring guides. Across arts disciplines, the responses of students who skipped a question or performance task (but who answered questions positioned later in the test booklet or in the *Creating/Performing* block) were scored as incorrect for unanswered multiple-choice questions or as Level 1 (Inadequate, Unsuccessful, or Unacceptable) for unanswered constructed-response items or performance tasks. A student's response for a given question was classified as not reached if a student failed to answer the question and all others following it in a given test booklet or *Creating/Performing* block. Data for students who did not reach a given question in a test booklet or in a *Creating/Performing* block were excluded from analysis for that question. Appendix C provides additional information on the processing of materials and item scoring.

For the Arts Assessment, more than 240,000 constructed responses were scored. This number includes rescoring to monitor interrater reliability. One hundred percent of student responses for Theatre *Performing* tasks were rescored in order to determine reliability rates. For all other types of questions and tasks across all arts disciplines, 25 percent of the student responses were rescored. The overall average percentages of exact agreement for the 1997 national reliability sample were 89.6 percent in Music, 84.6 percent in Theatre, and 86.2 percent in Visual Arts. Appendix D indicates the scoring reliability for each dichotomously or polytomously scored constructed-response item in the *Responding*, *Creating*, and *Performing* blocks in the 1997 Arts Assessment.

1.6 Reporting NAEP Arts Results

Responding results for Music, Theatre, and Visual Arts are summarized on three NAEP Item Response Theory (IRT) arts *Responding* scales, each of which ranges from 0 to 300. *Creating* and *Performing* results are not summarized on a standard NAEP IRT scale. To scale assessment results using IRT models, there must be a sufficient number of students taking a given group of exercises, and a sufficient number of exercises to be scaled of a given type. This was not the case for the *Creating* and *Performing* exercises in any of the three arts disciplines assessed. Although they consumed far more assessment time than written exercises, there were fewer exercises to group together into a scale. Moreover, given the complex administrative procedures associated with these tasks (such as videotaping responses, distributing arts materials, and having students work in groups), each student took only one such task. This prohibited the use of the type of IRT-scaling methodology used to summarize *Responding*

results. Instead of an IRT scale, *Creating* and *Performing* results are presented in terms of an average percent of the maximum possible score.

The Arts Assessment results cannot be reported in terms of the NAEP achievement levels (basic, proficient, and advanced). The complex, diverse nature of the assessment tasks in each arts discipline resulted in different scales for *Creating*, *Performing*, and *Responding*. Therefore, results could not be summarized for each arts discipline for the purpose of setting achievement levels.

1.7 Organization of the *NAEP 1997 Arts Technical Analysis Report*

Section 2 provides information about the special characteristics of the arts data, including reporting subgroups; student, school, and teacher samples; assessment instruments and questionnaires; sampling weights; and derived variables. *Sections 3, 4, and 5* include descriptions of the analyses used to summarize the results for the Arts Assessments in Music, Theatre, and Visual Arts, respectively. *Appendix A* describes the sample design and weighting procedures implemented. Field operations and data collection procedures are discussed in *Appendix B*. *Appendix C* provides information on the processing of assessment materials (e.g., printing of test booklets and questionnaires) and the scoring process. Scoring reliability data is listed in *Appendix D*. *Appendices E through G* provide lists of the items contributing to each scale, IRT parameters, and lists of items contributing to the average percents of the maximum possible scores.

Section 2

SPECIAL CONSIDERATIONS FOR THE NAEP 1997 ARTS ASSESSMENT ANALYSES³

*Alfred M. Rogers, Fred Yan, Terry L. Schoeps, and Debra M. Kline
Educational Testing Service*

Because of the special characteristics of the NAEP 1997 Arts Assessment, it is important for users to have an understanding of certain aspects of the assessment before examining the analysis of the data.

2.1 Student and School Samples

For the 1997 Arts Assessments in Music and Visual Arts, the national and regional data are based on nationally representative probability samples of all eighth-grade students. The results for Theatre are based on a representative sample of eighth-grade theatre students representing those schools in the nation with an extensive theatre curriculum. The samples were selected using a complex multistage sampling design that involved sampling students from selected schools within selected geographic areas across the country. The sample design had the following stages:

1. geographic areas were selected. (a county, group of counties, or metropolitan statistical area);
2. within each sampled geographic area, schools were selected (both public and nonpublic); and
3. within each sampled school, intact classrooms of students were selected for the Arts Assessments in Music and Visual Arts.

To ensure random sampling of students with training in each arts discipline, criteria for sampling classrooms specified that the subject taught in each classroom selected should not be the subject being assessed (e.g., classes sampled for the Arts Assessment in Visual Arts could not be visual arts classrooms; classes sampled for the Arts Assessment in Music could not be music classrooms). For the Arts Assessment in Theatre, students were sampled randomly from within eligible schools.

Each selected school that participated in the assessment and each student assessed represents a portion of the population of interest. Sampling weights are needed to make valid inferences between the student samples and the respective populations from which they were drawn. In addition, NAEP oversamples nonpublic schools and schools in which more than 15 percent of the student population's racial/ethnic makeup is other than White. Sampling weights adjust for disproportionate representation due

³ Nancy L. Allen edited Section 2 for use in this report.

to such oversampling. In the analysis of student data and reporting of results, nonresponse weighting adjustments have been made at both the school and student level, with the aim of making the sample of participating students as representative as possible of the entire eligible eighth-grade population. For details of the nonresponse weighting adjustment procedures, see the *NAEP 1998 Technical Report* (Allen, Donoghue, & Schoeps, 2001).

All eighth-grade students sampled for the Arts Assessment in Theatre were considered eligible if they had completed at least 30 classroom hours of instruction in theatre by the end of the 1996-97 school year. There were no course requirements for eligibility for the eighth-grade students sampled for the Arts Assessment in Visual Arts. In Music, there were no course requirements for eighth-grade students sampled to complete the blocks designed for the general student population. Students sampled for the supplementary (solo) *Creating* and *Performing* blocks in Music were selected from the sample of students initially selected from the general student population. Eligibility for the supplementary *Creating* and *Performing* blocks was limited to students who indicated that they either sang in a school chorus, sang outside of school, took singing lessons, played a musical instrument, took instrumental lessons, or played in a band or orchestra in school or outside of school. School officials advised assessment staff on SD/LEP accommodations necessary for individual students sampled for an assessment in a given arts discipline. Students for whom recommended SD/LEP accommodations could not be made were classified as ineligible for the assessment.

Table 2-1 contains, for public and nonpublic schools, the unweighted total number of participating schools, total number of students assessed, and the weighted school and student participation rates. In addition, Table 2-1 provides the weighted percentages of students who were matched to a completed school questionnaire. For Theatre, data indicating the overall teacher questionnaire participation rate and the total number of teachers completing a questionnaire are provided. In addition, the weighted percentage of Theatre students and total number of students matched to a completed Theatre teacher questionnaire are indicated.

The weighted school participation rates are calculated from the number of schools that were initially selected for the assessment. For each arts discipline, the numerator of this rate is the sum of the number of students represented by each initially selected school that participated in the assessment. The denominator is the sum of the number of students represented by each of the initially selected schools that had eligible students enrolled. The denominator included students in both participating and nonparticipating schools.

The weighted percentages of students who participated in the assessment reflect the weighted percentage of the eligible student population from participating schools within the jurisdiction. The numerator is the sum of the number of students represented by each student who participated in the assessment in either an initial session or a make-up session. The denominator of this rate is the sum, across all assessed students by arts discipline, of the number of students represented by each selected student who was eligible to participate, including students who did not participate.

The weighted school and student participation rates for the Theatre sample are lower than those typically obtained in NAEP, and, as is evident in Table 2-1, are substantially lower than those obtained for the Music and Visual Arts samples. As noted above, nonresponse adjustments were made to the

sampling weights used in the analysis of all three arts disciplines in order to compensate for school and student nonparticipation. However, the lower participation rates in the Arts Assessment in Theatre do raise questions about the effectiveness of these adjustments and the possibility of bias in the results.

Table 2-1
NAEP 1997 Arts Assessment
School, Student, and Teacher Questionnaire Participation Rates

	Weighted School Participation Rate Percentages	Total Number of Schools Participating	Weighted Student Participation Rate Percentages	Total Number of Students Assessed	Weighted Percentage of Students Matched to a School Questionnaire	Total Number of Students Matched to a School Questionnaire
Music	80	98	91	2,275	94	2,114
<i>Public</i>	79	84	91	1,999		
<i>Nonpublic</i>	83	14	94	276		
Theatre	67	42	82	1,386	92	1,193
<i>Public</i>	69	40	79	1,335		
<i>Nonpublic</i> ¹	40	2	93	51		
Visual Arts	84	128	91	2,999	93	2,799
<i>Public</i>	84	116	90	2,756		
<i>Nonpublic</i>	85	12	95	243		
	Teacher Questionnaire Participation Rate Percentage		Total Number of Teachers Completing a Questionnaire		Weighted Percentage of Students Matched to a Teacher Questionnaire	Total Number of Students Matched to a Teacher Questionnaire
Theatre	95		57		98	1,320

¹ Results were not reported for nonpublic-school students in the Theatre Assessment.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

In order to provide some evidence on the extent of potential bias, analyses were conducted comparing the participating Theatre schools to all schools in the Theatre sample with respect to five demographic characteristics:

- metro status (metro, nonmetro);
- type of locale (large city, midsize city, urban fringe of large city, urban fringe of midsize city, large town, small town, rural);
- affiliation (public, nonpublic),
- school type (public, Catholic, other religious, nonsectarian); and
- region (Northeast, Southeast, Central, West).

In addition, comparisons were made for three additional school enrollment variables: grade 8 enrollment, total school enrollment, and percentage of minority enrollment. In general, the participating schools were similar to the full sample of schools for most of the demographic and school enrollment variables. The participating schools were, however, more likely to come from large city locales and had, on average, lower grade 8 and total enrollments than nonparticipating schools.

2.2 THEATRE TEACHER RESPONSE DATA

Theatre teachers of the students participating in the NAEP Arts Assessment in Theatre were asked to complete a questionnaire about instructional practices, general teaching background, and theatre teaching background. The questionnaire was divided into three parts. The first part contained 85 questions about teachers' general educational background and training. The second part contained 25 questions pertaining to teachers' background, activities, and preparation in theatre. The third section contained 50 questions on specific instructional practices. A summary of teacher participation and students matched to teachers is included in Table 2-1. The purpose of the teacher sample is to estimate the numbers of students whose teachers have various attributes, not to estimate the attributes of the teacher population.

2.3 SCHOOL RESPONSE DATA

Principals of students sampled for the assessment were asked to complete a questionnaire about the school's characteristics and students' access to instruction in the arts. The school questionnaire covered three broad areas. The first part pertained to the availability of courses in the arts. It also contained questions about students' access to computers because of an interest in planning for a future NAEP research program in the area of computer testing. The second part asked questions about the status of staff members teaching in the arts, the facilities and available resources for the arts, and the existence of special programs in the arts, such as artists-in-residence and summer arts programs. The final part of the school questionnaire pertained to demographics at the school, such as school enrollment. It also included variables used to describe the general climate of the school, such as attendance rates of students and staff, and the frequency of various problems in the school. The purpose of the school sample is to estimate the numbers of students whose schools have various attributes, not to estimate the attributes of schools.

2.4 SD/LEP DATA

An additional questionnaire was designed to gather information about students with disabilities (SD) and limited-English proficient (LEP) students. The SD/LEP questionnaire was completed by a school staff member knowledgeable about those students who were selected to participate in the assessment and who were identified as (1) having an Individual Education Plan (IEP) or equivalent plan (for reasons other than being gifted and talented) or (2) being limited-English proficient. A questionnaire was completed for sampled students who were disabled or limited-English proficient regardless of whether the student participated in the assessment. Each questionnaire took approximately five minutes to complete and asked about the student's background and the special programs in which he or she may have been enrolled.

2.5 USING WEIGHTS IN ANALYSES OF NAEP ARTS DATA

Following the collection of assessment and background data from and about assessed students, sampling weights and associated sets of replicate weights were derived. The sampling weights are needed to make valid inferences from the student samples to the respective populations from which they were drawn and should be used for all analyses, whether exploratory or confirmatory. Replicate weights are used in the estimation of sampling variance, through the procedure known as jackknife repeated replication.

In the NAEP sampling scheme, students do not have an equal probability of being selected. Therefore, as in all complex surveys, each student has been assigned a sampling weight. The larger the probability of selection for students within a particular demographic group, the smaller the weights for those students will be. When computing descriptive statistics or conducting inferential procedures, one should weight the data for each student. Performance of statistical analyses without weights can lead to misleading results.

Another way in which the complex sample design used by NAEP differs from simple random sampling is that the NAEP sampling scheme involves the selection of clusters of students from the same school, as well as clusters of schools from the same geographically defined primary sampling unit, or PSU. As a result, observations are not independent of one another as they are in a simple random sample. Therefore, use of standard formulas for estimating the standard error of sample statistics such as means, proportions, or regression coefficients will result in values that are generally too small. The standard error, which is a measure of the variability of a sample statistic, gives an indication of how well that statistic estimates the corresponding population value. It is used to conduct tests of statistical significance. If conventional simple random sampling formulas are used to compute standard errors, too many statistically significant results will occur in most instances.

The following describes the weight variables on the data files and provides guidelines for their use. Further details of the weighting procedures are given in the *NAEP 1998 Technical Report* (Allen, Donoghue, & Schoeps, 2001); further information is provided in Appendix A.

2.5.1 Overall Weights

Each student was assigned a weight to be used for making inferences about the nation's students. This weight is known as the overall weight. The overall weight contained three components—a base weight, an adjustment for school nonparticipation, and an adjustment for student nonparticipation.

The base weight assigned to a student is the reciprocal of the probability that the student was selected for a particular assessment. That probability is the product of four factors:

- the probability that the PSU was selected;
- the conditional probability, given the PSU, that the school was selected;
- the conditional probability, given the sample of schools in a PSU, that the school was allocated the specified assessment; and

- the conditional probability, given the school, that the student was selected for the specified assessment.

The base weight for a selected student was adjusted by two nonresponse factors. The first of these was to adjust for sessions that were not conducted, either because a contacted school refused or because a cooperating school did not conduct all of its assigned sessions. The second adjustment was needed to adjust for students who were (or should have been) invited to the assessment but did not appear either in the scheduled session or a makeup session. Specific information about the weights used for the *Responding*, *Creating*, and *Performing* scales for each arts discipline are available in *The NAEP 1997 Arts Data Companion* (Rogers, Yan, Schoeps, & Kline, 2000).

2.5.2 Replicate Weights

In addition to estimation weights, a set of replicate weights was provided for each student. Each overall weight has a set of corresponding replicate weights, which are used for estimating the sampling errors of estimates derived using the overall weights. These weights are designed to reflect the method of sampling schools, and account for the type of stratification used and whether or not the student's school was included in the sample with certainty. The method of sampling students within schools is also reflected, implicitly in the case of noncertainty schools and explicitly for schools included with certainty. These overall replicate weights also reflect the impact on sampling errors of the student-level nonresponse adjustment applied to the overall weights.

These replicate weights are used in calculating the sampling errors of estimates obtained from the data, using the jackknife repeated replication method. The methods of deriving these weights were aimed at reflecting the features of the sample design appropriately, so that when the jackknife variance estimation procedure is implemented as intended, approximately unbiased estimates of sampling variance result.

The jackknife variance estimation procedure is described in Section 10.5 of the *NAEP 1998 Technical Report* (Allen, Donoghue, & Schoeps, 2001). The variance of the full sample is estimated by jackknife repeated replication. This process involves repeatedly selecting portions of the sample to calculate the statistic of interest. The estimates that result are called replicate estimates. The variability among these calculated quantities is used to obtain the full sample variance. The process of forming these replicate estimates involves first dividing the sample elements among a set of replicate groups, then using the pattern of replicate groups in a systematic fashion to apply replicate weights to the file.

Table 2-2 gives the number of students and the sum of their weights for each student sample in the 1997 Arts Assessment.

Table 2-2
NAEP 1997 Arts Assessment
Number of Assessed Students by Sample and Item Type

	<i>Responding Items</i>		<i>Creating/Performing Items</i>		<i>Supplementary (Solo) Creating and Performing Items</i>	
	Total	Sum of Weights	Total	Sum of Weights	Total	Sum of Weights
Music	2,275	3,568,739	1,989	10,707,952	567	1,969,932
Theatre	1,386	125,758	1,225	349,376	†	†
Visual Arts	2,999	3,590,804	2,129	10,764,019	†	†

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

2.6 STUDENT ASSESSMENT INSTRUMENTS

2.6.1 Music, Theatre, and Visual Arts Exercises

The assessments in Music, Theatre, and Visual Arts included “blocks” or sets of questions, of approximately 25 or 50 minutes. (Blocks are collections of questions grouped, in part, according to the amount of time required to answer them.) Each block consisted of one or more stimuli and sets of multiple-choice, constructed-response, or *Creating/Performing* items to assess students’ mastery of material. Table 2-3 summarizes the number of blocks by arts discipline.

Table 2-3
NAEP 1997 Arts Assessment
Total Number of Blocks by Arts Discipline

	Music		Theatre		Visual Arts	
	Total	Number Completed by an Individual Student	Total	Number Completed by an Individual Student	Total	Number Completed by an Individual Student
<i>Creating/Performing</i>	5 ¹	1 or 2 ²	3	1	3	0 or 1
<i>Responding</i>	4	2	4	2	4	1 or 2

¹ The five *Creating/Performing* blocks in the Arts Assessment in Music included three blocks for the general student population and two blocks for a subsample of students who were currently enrolled in a music activity.

² Students who were selected as part of the subsample of Music students completed two blocks: one block of activities for students from the general population and one block of activities for students in the Music subsample.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The grade 8 Arts Assessment in Music included five *Creating* and *Performing* blocks and four *Responding* blocks (also called A/B blocks). The five *Creating* and *Performing* blocks were divided into three *Creating* and *Performing* blocks for students in the general population, and two *Creating* and

Performing blocks for students currently active in some type of music activity. All students sampled for the Arts Assessment in Music completed one of the *Creating* and *Performing* blocks for the general student population and two *Responding* blocks. In addition, a small sample of students who indicated current involvement in a music activity completed one of the two additional *Creating* and *Performing* blocks as a fourth block.

The grade 8 Arts Assessment in Theatre included three *Creating/Performing* blocks and four *Responding* blocks. All Theatre students completed one *Creating/Performing* block and two *Responding* blocks.

The grade 8 Arts Assessment in Visual Arts included three *Creating* blocks and four *Responding* blocks. All students sampled for Visual Arts completed either one *Responding* block and one *Creating* block or two *Responding* blocks. In Visual Arts, three of the four *Responding* blocks featured *Creating* tasks.

The data presented in Table 2-4 reflect the number of questions by item type for the 1997 Arts Assessment. The assessment pool for the three arts disciplines assessed contained a total of 168 unique questions—41 multiple-choice, 97 constructed-response, and 30 *Creating/Performing* tasks. Some of the Arts Assessment test questions were scored several times to evaluate different aspects of the responses. When numbers of items are specified in the rest of this report, they include these separate scorings of responses as separate items. Using information gathered from the field test, the booklets were carefully constructed to balance time requirements for the questions in each block. Tables 2-5, 2-6, and 2-7 show how the blocks of exercises were formed into booklets for each arts discipline.

Table 2-4
NAEP 1997 Arts Assessment
Distribution of Questions by Item Type

	Music	Theatre	Visual Arts
Multiple-Choice	21	8	12
Constructed-Response	35	38	24
<i>Performing/Creating</i>	14	6	10

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table 2-5
NAEP 1997 Arts Book Map
Music

Booklet Number	Responding Blocks		Creating and Performing Blocks	Solo Blocks	Target Sample Size
21	C	D	G	J	250
22	E	C	H	K	250
23	C	F	I	J	250
24	D	E	I	K	250
25	F	D	H	J	250
26	E	F	G	K	250
27	D	C	G	J	167
28	D	C	H	K	83
29	F	E	H	J	83
30	F	E	I	K	167

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table 2-6
NAEP 1997 Arts Book Map
Theatre

Booklet Number	Responding Blocks		Creating/ Performing Blocks	Target Sample Size
31	C	D	G	250
32	E	C	H	250
33	C	F	I	250
34	D	E	I	250
35	F	D	H	250
36	E	F	G	250
37	D	C	G	167
38	D	C	H	83
39	F	E	H	83
40	F	E	I	167

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table 2-7
NAEP 1997 Arts Book Map
Visual Arts

Booklet Number	Responding Blocks		Creating Blocks	Target Sample Size
1	E	C	—	100
2	C	D	—	100
3	D	F	—	100
4	D	E	—	100
5	F	C	—	100
6	E	F	—	100
7	D	C	—	100
8	F	E	—	100
9	E	—	G	140
10	E	—	H	140
11	E	—	I	140
12	C	—	G	140
13	C	—	H	140
14	C	—	I	140
15	D	—	G	140
16	D	—	H	140
17	D	—	I	140
18	F	—	G	140
19	F	—	H	140
20	F	—	I	140

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

2.6.2 Background Questionnaires

Each booklet in the Assessment also included several sets of background questionnaires. Students sampled for the 1997 Arts Assessment completed one 5-minute set of student demographic background questions and one 10-minute set of subject-specific background questions. The subject-specific background questionnaires were designed to gather contextual information about students, their instructional and out-of-school arts experiences, and their attitudes toward the arts discipline in which they were being assessed.

The student demographic background questionnaire included a common set of questions about students' race/ethnicity, language spoken in the home, parents' level of education, number of reading materials in the home, amount of television watched by students, the amount of homework assigned, and which parents live at home.

The student subject-specific questionnaires covered three general categories of information: students' interest in the subject; students' in-school experiences in the subject; and students' out-of-school experiences in the subject. These 10-minute sections contained 37 questions for Music, 45 questions for Theatre, and 33 questions for Visual Arts.

The types of questions asked in each of the three categories of the subject-specific questionnaires are described as follows: students' interest in the subject included students' ratings of their interest and ability in the subject. For example, in Music, three of the statements to which students were asked to respond "Agree," "Not Sure," or "Disagree," included: "I like to listen to music," "I think I have talent for music," and "People tell me I am a good musician." Students' in-school experiences were characterized by the frequency with which their teachers provided various subject-related instructional activities during class and by student participation in various arts-related activities during school. Students' out-of school experiences were characterized by the frequency with which students were involved in various arts-related activities outside of school, not in connection with school work.

2.7 NAEP REPORTING GROUPS

Results for the 1997 assessment were calculated for student subgroups defined by gender, race/ethnicity, parents' education, school type, region of the country, type of location, and *Responding* score groups. The following explains how each of these subgroups was derived.

DSEX (Gender)

The variable SEX is the gender of the student being assessed, as taken from school records. For a few students, data for this variable was missing and was imputed by ETS after the assessment. The resulting variable DSEX contains a value for every student and is used for gender comparisons among students.

DRACE (Race/Ethnicity)

The variable DRACE is an imputed definition of race/ethnicity, derived from up to three sources of information. This variable is used to compare the performance of race/ethnicity subgroups. Two questions from the set of general student background questions were used to determine race/ethnicity.

If you are Hispanic, what is your Hispanic background?

- I am not Hispanic.
- Mexican, Mexican American, or Chicano
- Puerto Rican
- Cuban
- Other Spanish or Hispanic background

For the question above, students who responded by filling in the second, third, fourth, or fifth oval were considered Hispanic. For students who filled in the first oval, did not respond to the question, or provided information that could not be classified, responses to the following question were examined to determine their race/ethnicity.

Which best describes you?

- White (not Hispanic)
- Black (not Hispanic)
- Hispanic (“Hispanic” means someone who is Mexican, Mexican American, Chicano, Puerto Rican, Cuban, or from some other Spanish or Hispanic background.)
- Asian or Pacific Islander (“Asian or Pacific Islander” means someone who is Chinese, Japanese, Korean, Filipino, Vietnamese, or from some other Asian or Pacific Island background.)
- American Indian or Alaskan Native (“American Indian or Alaskan Native” means someone who is from one of the American Indian tribes, or one of the original people of Alaska.)
- Other (What?) _____

Students’ race/ethnicity was then assigned on the basis of their responses. For the question above, students who filled in the sixth oval (“Other”), or who did not respond at all, race/ethnicity was assigned as determined by school records. Race/ethnicity could not be determined for students who did not respond to either of the demographic questions and whose schools did not provide information about race/ethnicity.

PARED (Parents’ Highest Level of Education)

The variable PARED is derived from responses to two questions, B000005 and B000006, from the set of general student background questions. Students were asked to indicate the extent of their mother’s education in question B000005:

How much education did your mother receive? (“Mother” can be a mother, stepmother, or female guardian.)

- She did not finish high school.
- She graduated from high school.
- She had some education after high school.
- She graduated from college.

Students were asked a similar question about their father’s education level (B000006):

How much education did your father receive? (“Father” can be a father, stepfather, or male guardian.)

- He did not finish high school.
- He graduated from high school.
- He had some education after high school.
- He graduated from college.

The information was combined into one parental education reporting variable through the following process. If a student indicated the extent of education for only one parent, that level was included in the data. If a student indicated the extent of education for both parents, the higher of the two levels was included in the data. If the student did not respond for either parent, the student was recorded as having provided no response.

SCHTYP7 (Type of School—Music and Visual Arts Only)

The variable SCHATYP7 is provided by Westat, and is used to determine the type of school that a student attended. The values for this variable are as follow:

- 1 Public
- 2 Catholic
- 3 Other religious
- 4 Unaffiliated
- 5 Other/nonreligious
- 6 Department of Defense Education Activity (DoDEA)
- 7 Bureau of Indian Affairs

NAEPRGN (Region of the Country)

Results are reported for four regions of the nation: Northeast, Southeast, Central, and West. States included in each region are shown in the following list.

Northeast: Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia (DC metropolitan statistical area only)

Southeast: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia (other than DC metro area), West Virginia

Central: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin

West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, Wyoming

TOL3 (Type of Location)

Results are provided for students attending schools in three mutually exclusive location types—central city, urban fringe/large town, and rural/small town—as defined below. The type of location variable is defined in such a way as to indicate the geographical location of a student's school. The intention is not to indicate, or imply, social or economic meanings for these location types.

Central City: The Central City category includes central cities of all metropolitan statistical areas (MSAs). Central City is a geographic term and is not synonymous with "inner city."

Urban Fringe/Large Town: An Urban Fringe includes all densely settled places and areas within MSAs that are classified as urban by the U.S. Census Bureau. A Large Town is defined as places outside MSAs with a population greater than or equal to 25,000.

Rural/Small Town: Rural includes all places and areas with a population of less than 2,500 that are classified as rural by the U.S. Census Bureau. A Small Town is defined as places outside MSAs with a population of less than 25,000 but greater than or equal to 2,500.

Responding Score Groups

Low-, middle-, and high-scoring groups were formed by sorting the *Responding* scale scores from lowest to highest, calculating cumulative percentiles, and partitioning at the 25th and 75th percentiles.

2.8 DERIVED VARIABLES

Several variables on the arts data files were derived from the systematic combination of response values for one or more items from either the student background questionnaires or the Theatre teacher questionnaire.

Three variables derived from the student Theatre background questionnaire and three variables from the Theatre teacher questionnaire were created as the sum of "yes" responses to a series of background questions. Variable BTDV0A9 was derived from the sum of "yes" responses to student background items about in-school activities in which the student participated:

Questions 9-13. If you have been involved in live theatre productions **in school**, which of the following activities have you done? Fill in **all** ovals that apply.

- | | |
|---|-----------------------|
| 9. Acting | <input type="radio"/> |
| 10. Technical work (such as costumes, makeup, lighting, scenery, props) | <input type="radio"/> |
| 11. Directing | <input type="radio"/> |
| 12. Writing a play or script | <input type="radio"/> |
| 13. I have not been involved. | <input type="radio"/> |

Variable BTDV028 was derived from the sum of “yes” responses to student background items about the student’s participation in any film or video productions. Variable BTDV037 was derived from the sum of “yes” responses to student background items about the student’s participation in extracurricular activities in theatrical performances (e.g., plays, musicals, puppet shows). Variable TPDV0A8 was derived from the sum of “yes” responses to teacher questionnaire items about the teacher’s professional development in the past five years. Variable TCDVA12 was derived from the sum of “yes” responses to teacher questionnaire items about the school’s stage facilities available for theatrical productions. Variable TCDV0A2 was derived from the sum of “yes” responses to teacher questionnaire items about the number of curriculum types the teacher used in theatre instruction at the school.

Teacher variables TPDV008–TPDV014 were created from items TP00008–TP00014, which asked teachers to check all responses that applied to questions about professional development. The responses were recategorized as follows:

- 1 Yes (I did this)
- 2 No (I did not do this, but I did something else among these professional development activities)
- 3 No (I did none of these professional development activities)

2.9 NAEP ARTS SCALES AND SCORES

For each arts discipline, analyses were conducted to determine the percentage of students who gave various responses to each cognitive and background question. Item response theory (IRT) was used to estimate average proficiency for the nation and various subgroups of interest within the nation for items classified as *Responding* items. Mean percent-correct scores were developed for items classified as *Creating*, *Performing*, or *Creating/Performing*.

2.9.1 NAEP Arts *Responding* Scales

A primary method by which results from the 1997 Arts Assessment were disseminated is scale-score reporting. With scaling methods, the performance of a sample of students in an arts discipline can be summarized on a single scale even when different students have been administered different items.

NAEP uses the methodology of marginal estimation of group score distributions to estimate characteristics of the proficiency distributions.

IRT models the probability of answering a question correctly as a mathematical function of proficiency or skill. The main purpose of IRT analysis is to provide a common scale on which performance can be compared across groups, such as those defined by age, assessment year, or subpopulations (e.g., race/ethnicity or gender) and to make it possible to assess groups on more material than would be practical to give to any one student.

In the NAEP design, students do not receive enough questions about a specific topic to permit reliable estimates of individual performance. Traditional test scores for individual students, even those based on IRT, would contribute to misleading estimates of population characteristics, such as subgroup averages and percentages of students at or above a certain proficiency level. Instead, NAEP uses marginal maximum likelihood estimation of group score distributions to estimate the distribution of proficiency in the population. NAEP does not produce scale scores for individuals. Statistics describing performance of groups of students on the NAEP scales are calculated. These statistics estimate values that would have been obtained had individual proficiencies been observed exactly; that is, had each student responded to a sufficient number of cognitive questions so that his or her proficiency could be precisely estimated.

For the Music, Theatre, and Visual Arts *Responding* exercises, separate IRT *Responding* scales were constructed. Because of the separate scaling, each of the three arts disciplines employed slightly different steps in data analysis and IRT scaling. Because of concerns about multidimensionality, and because there were too few items to create IRT scales for *Creating*, *Performing*, or *Creating/Performing*, only items in the *Responding* categories of the arts framework were entered into the IRT scaling procedure. A single IRT scale was created for each of the three arts disciplines. (In Music, it was necessary to fit two related scales and then combine them into a composite. The two interim Music subscales differed in the dependence of their tasks on music notation and technical vocabulary.) The IRT parameters on which the scales were based are provided in Appendix F.

The *Responding* IRT scales were created with a mean of 150 and a standard deviation of 35; the great majority of students had scores between 45 and 255. Although the mean of the *Responding* scale for each subject was set to 150, the scales measure different accomplishments. Comparisons cannot be made between student results on any pair of *Responding* scales, even though the scales share the same mean (150). In other words, a score of 165 in Visual Arts is not necessarily “better” than a score of 160 in Music.

2.9.2 NAEP Arts *Creating* and *Performing* Scores

Because there were too few items to create IRT scales for *Creating* items in Visual Arts and Music, *Creating/Performing* items in Theatre, and *Performing* items in Music, these were formed into separate percent-of-total-possible-points averages, with mean percent-correct scores reported at various levels. Certain Theatre items which combined aspects of *Responding* and *Creating* (e.g., “draw a set design for this play”) did not fit the *Responding* IRT scale, and so were not included in that scale, but were reported on an item-level basis. All of the *Creating* and *Performing* tasks that were administered to a subsample of students in the Arts Assessment in Music were also reported on an item-level basis. In

addition, items with logical dependencies (e.g., write a new ending to a script, followed by a discussion on how that ending accomplished one's goals) were separated, with the discussions included in the IRT scale. The endings on which they depended were put in the *Creating/Performing* averages. The scales are listed in Table 2-8.

Table 2-8
NAEP 1997 Arts Score Scales

Scale	Type of Scale
Music <i>Responding</i> Composite Scale	Composite of 2 IRT scales
Music <i>Performing</i> Scale	Mean-percent correct
Music <i>Creating</i> Scale	Mean-percent correct
Theatre <i>Responding</i> Scale	IRT
Theatre <i>Creating/Performing</i> Scale	Mean-percent correct
Visual Arts <i>Responding</i> Scale	IRT
Visual Arts <i>Creating</i> Scale	Mean-percent correct

NOTE: Music Supplementary (Solo) *Performing* and *Creating* items were not placed into scales, but results are available for each item for the self-selected samples of students responding to each item.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

2.10 DRAWING INFERENCES FROM THE RESULTS

The use of confidence intervals, based on the standard errors, provides a way to make inferences about the population averages and percentages in a manner that reflects the uncertainty associated with the sample estimates. An estimated sample scale score average ± 2 standard errors represents about a 95-percent confidence interval for the corresponding population size. This means that with 95-percent certainty, the average performance of the entire population of interest is within about ± 2 standard errors of the sample average.

As an example, suppose that the average Theatre *Responding* scale score of students in a particular group was 256, with a standard error of 1.2. A 95-percent confidence interval for the population quantity would be as follows:

$$\begin{aligned}
 \text{Average} \pm 2 \text{ standard errors} &= 256 \pm (2 \times 1.2) \\
 &= 256 \pm 2.4 \\
 &= (256 - 2.4, 256 + 2.4) \\
 &= (253.6, 258.4)
 \end{aligned}$$

Thus, one can conclude with close to 95-percent certainty that the average scale score for the entire population of students in that group is between 253.6 and 258.4.

Similar confidence intervals can be constructed for percentages, provided that the percentages are not extremely large or extremely small. For percentages, confidence intervals constructed in the above manner work best when sample sizes are large, and the percentages being tested have magnitude relatively close to 50 percent. Statements about group differences should be interpreted with caution if at least one of the groups being compared is small in size and/or if “extreme” percentages are being compared. For reporting purposes, groups’ percentages, P , were treated as “extreme” if

$$P < P_{\text{lim}} = \frac{200}{N_{\text{EFF}} + 2}$$

where the effective sample size,

$$N_{\text{EFF}} = \frac{P(100 - P)}{(SE_{JK})^2}$$

and SE_{JK} is the jackknife standard error of P . This “rule of thumb” cut-off leads to flagging a large proportion of confidence intervals that would otherwise include values < 0 or > 1 . Similarly, at the other end of the 0–100 scale, a percentage is deemed extreme if $(100 - P) < P_{\text{lim}}$. In either extreme case, the confidence intervals described above are not appropriate, and procedures for obtaining accurate confidence intervals are quite complicated. In this case, the value of P was reported, but no standard error was estimated and hence no tests were conducted. This rule was not applied to response percentages within items.

As for percentages, confidence intervals for average scale scores are most accurate when sample sizes are large. For some of the subgroups of students for which average scale scores or percentages were reported, student sample sizes could be quite small. For results to be reported for any subgroup, a minimum student sample size of 62 was required. If students in a particular subgroup were clustered within a small number of geographic primary sampling units (PSUs), the estimates of the standard errors might also be inaccurate. So, subgroup data were required to come from a minimum of five PSUs.

2.10.1 Analyzing Group Differences in Averages and Percentages

To determine whether there is a real difference between the average scale score (or percentage of a certain attribute) for two groups in the population, one needs to obtain an estimate of the degree of uncertainty associated with the difference between the average scale scores or percentages of these groups for the sample. This estimate of the degree of uncertainty, called the standard error of the difference between the groups, is obtained by squaring each group’s standard error, summing these squared standard errors, and then taking the square root of this sum.

If zero is within the confidence interval for the differences there is no statistically significant difference between the groups. The multiplier is the .975 (1–.025) percentile from a t -distribution with the degrees of freedom that vary by the values of the average scale scores, their standard errors, and the number of PSUs that contribute to the average scale scores. See the *NAEP 1998 Technical Report* (Allen, Donoghue, & Schoeps, 2001) for more details. This procedure produces a conservative estimate of the standard error of the difference, since the estimates of the group averages or percentages will be

positively correlated to an unknown extent due to the sampling plan. Direct estimation of the standard errors of all reported differences would involve a heavy computational burden.

Sometimes a group of related comparisons are made, such as comparing the average scale scores for a particular region with those from another region for specific groups of students. If one wants to hold the certainty level for a specific set of comparisons at a particular level (e.g., 95 percent), adjustments (called multiple-comparisons procedures) need to be made. One such procedure, the Bonferroni method, was used to form confidence intervals for the differences for sets of comparisons. The set of comparisons is referred to as a “family,” and the typical family involves all subgroups related by a certain background question (Miller, 1996).

Multiple-comparisons procedures, like the Bonferroni method, are useful for controlling the overall Type I error rate for a defined set of hypothesis tests. However, especially when the number of potential comparisons which could be made is large, as in NAEP data, this protection comes at the substantial loss of power in detecting specific consistent patterns in the data. For example, more powerful and complex tests of significance designed to identify consistent patterns in the data might judge that two groups were significantly different when a Bonferroni multiple-comparisons procedure would not.

These tests do not control the overall Type I error rate when they are applied to several related subgroups, such as the students in each region of the country. For this reason, the Bonferroni method for controlling Type I error was used when related subgroups were tested. For example, when tests were conducted for means or percentages for the separate race/ethnicity groups (e.g., White, Black, and Hispanic) these tests were treated as a single family of comparisons of size 3. The significance level for each of the separate tests was adjusted by the Bonferroni procedure to yield a Type I error rate of .05 for the family of comparisons.

2.10.2 Minimum Sample Sizes Used For Reporting in the *NAEP 1997 Arts Report Card*

Because of the relatively small sample sizes in the Arts Assessment, the number of students in some of the demographic subpopulations was not sufficiently high to permit accurate estimation of performance and/or background variable results. As a result, data are not provided in *The NAEP 1997 Arts Report Card* (Persky, Sandene, & Askew, 1998) for subgroups with students from very few schools or for subgroups with very small sample sizes. For results to be reported for any subgroup, at least five PSUs had to be represented in the subgroup. In addition, a minimum sample of 62 students per subgroup was required. For statistical tests pertaining to subgroups, the sample size for both groups had to meet the minimum sample size requirements.

Section 3

DATA ANALYSIS FOR THE 1997 ARTS ASSESSMENT IN MUSIC⁴

*Frank Jenkins, Spencer S. Swinton, Laura J. Jerry, Edward Kulick, and Shuyi Hua
Educational Testing Service*

3.1 INTRODUCTION

This section describes the analyses of the cognitive and background item responses in the NAEP 1997 Arts Assessment in Music at grade 8. This section focuses on the methods and procedures used to estimate distributions of IRT-based scale scores⁵ and percent-correct summaries for subgroups of students. The wide array of topics include, for example, the scoring of constructed-response items, classical item statistics, mean item scores, item response theory (IRT) analysis of the Music scale, and estimation of subgroup means through marginal maximum likelihood estimation of group score distributions. The statistical bases of the IRT and marginal maximum likelihood methodology used in NAEP and described in this section are given in the *NAEP 1996 Technical Report* (Allen, Carlson, & Zelenak, 1999) and the *NAEP 1998 Technical Report* (Allen, Donoghue, & Schoeps, 2001). These analyses serve as a basis for the results presented in the Music sections of the *NAEP Arts Report Card* (Persky, Sandene, & Askew, 1998).

There were two major steps in the analysis of the NAEP 1997 Music data, each of which is described in a separate section:

- creating the item response theory (IRT) scales for the *Responding* items (see Section 3.2) and
- calculating the mean percent-correct scales for the *Creating* items (see Section 3.3).

3.2 CREATING THE ITEM RESPONSE THEORY (IRT) SCALE FOR MUSIC *RESPONDING* ITEMS

3.2.1 Item Analysis

This section contains a description of the item analysis performed using nationally representative sample data. The analysis examines items within blocks. In preparation for this step, constructed-response items were scored by persons trained as raters, and derived background variables were calculated. Item statistics such as mean item score, average block score, item to total score correlations and percent responding in each item category were calculated.

⁴ Spencer S. Swinton was primarily responsible for the planning, specification, and coordination of the Music analyses. Computing activities for all Music scaling and data analyses were directed by Edward Kulick and completed by Laura J. Jerry and Shuyi Hua. Others contributing to the analysis of Music data were David S. Freund, Bruce A. Kaplan, and Katharine E. Pashley.

⁵ Scale scores based on item response theory (IRT)

In NAEP analyses (see Section 2) a distinction is made between missing responses at the end of each block (not reached) and missing responses prior to the last completed response (omitted). Not-reached items are those occurring after the last item a student answers in a block. Items that were not reached were treated as if they had not been presented to the student, and were therefore not counted as wrong. An omission occurs when a student skips an item but responds to some item(s) after it. Omitted items are scored as incorrect. The proportion of students attempting the last item of a block is often used as an index of the degree of speededness of the block of items (i.e. the smaller the proportion of students reaching the end of the block, the more speeded the block).

Table 3-1 shows block statistics for *Responding* items, that is, standard multiple-choice or constructed-response items presented in a paper and pencil format, which were analyzed with IRT methods. In contrast, *Creating* and *Performing* items are those which require participation in a musical or composing performance task (see Section 3.3). The table shows the number of scaled items, number of scaled constructed-response items, number of scaled multiple-choice items, sample size, mean item score, alpha reliability, mean item to total score correlation, and the proportion of students attempting the last item in the block for each block. These values were calculated within block only for those items used in the scaling process. Student sample weights were used to calculate statistics, except for the sample size. The results indicated that the blocks differ in number of items, average difficulty, reliability, and percent reaching the last item. ME is the easiest block (mean item score = .63) with the highest reliability (alpha = .77 and average r-Polyserial = .62). In contrast, block MF is the most difficult (mean item score = .31), with the lowest reliability (alpha = .63). None of these blocks are extremely speeded.

Table 3-1
NAEP 1997 Arts Assessment Descriptive Statistics
for the IRT-Scaled Responding Items by Block for the Music Sample¹

Statistic	MC	MD	ME	MF
Number of Items Before Clustering	29	20	25	35
Number of Items After Clustering	15	17	17	20
Number of Scaled Items	15	17	14	10
Number of Scaled Constructed-Response Items	10	10	9	7
Number of Scaled Multiple-Choice Items	5	7	5	3
Unweighted Sample Size	1,196	1,113	1,115	1,111
Weighted Mean Item Score Averaged Over Block	.39	.48	.63	.31
Weighted Alpha Reliability	.75	.72	.77	.63
Weighted Average r-Polyserial	.60	.55	.62	.57
Weighted Proportion of Students Attempting Last Item	.87	.99	.93	.93

¹ The *Performing* items from block ME are not described here. Some of the Arts Assessment items were scored several times to evaluate different aspects of responses. The items specified in this table include these separate scorings of responses as separate items.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Approximately 25 percent of the responses to constructed-response items were scored twice to assess interrater reliability. Appendix D lists the reliability information. Dichotomous constructed-

response items were scored on a scale of 1 to 2, while polytomous constructed-response items were scored on a scale of 1 to 3, 1 to 4, or 1 to 5 to reflect degrees of knowledge. (In IRT parameter estimation, this scale is shifted down by 1, as indicated in Table 3-2.) The reliability for dichotomized responses was in an acceptable range comparable to other assessments, with proportion exact agreement ranging from .90 to 1.00 (averaging .97). The Cohen's Kappa statistic, which adjusts reliability estimates of dichotomous responses for chance agreement, ranged from .82 to .99 (averaging .95), as shown in Table D-1 in Appendix D. Table D-2 indicates that for the polytomous items, proportion exact agreement was in a range comparable to other assessments with values from .84 to .99 (averaging .91). The intraclass correlation (appropriate for polytomous items) is a reliability estimate that corrects for chance agreement. These also were in an acceptable range with values from .72 to 1.00 and an average of .91.

3.2.2 Estimation of Item Parameters

For each arts discipline (Music, Theatre, and Visual Arts), separate IRT scales were constructed using the NAEP BILOG/PARSCALE program (Mislevy & Bock, 1982; Muraki & Bock, 1991). Two IRT scales were constructed for Music, denoting dependence on music notation and technical vocabulary. The program uses marginal estimation procedures to estimate the parameters of the one-, two-, and three-parameter logistic models, and the generalized partial credit model (Muraki, 1992). Dichotomous multiple-choice items were analyzed using a three-parameter IRT model; dichotomized constructed-response items employed a two-parameter IRT model. Three-, four-, and five-category items were polytomously scored and were analyzed with a generalized partial credit model (Muraki, 1992).

Creating and *Performing* items were written to different specifications than *Responding* items and evidenced dependencies that would violate the local independence assumption of IRT. As a result, they were not included in the IRT scale, but were used to form a separate scale based on a percent-correct score (see Section 3.3).

Item responses that were missing prior to the last completed item in a block were considered omitted and scored as wrong. Also, items that were not reached were treated as if they were not presented to the student (and not counted as wrong). For IRT scaling, omitted multiple-choice items were treated as fractionally correct ($1 \div$ [number of alternatives]). Responses to constructed-response items that were off-task were treated as omitted and assigned to the lowest category. For Music, constructed-response items had two, three, four, and five categories of increasing credit. Table 3-2 lists the score level definitions for constructed-response items which were scored with partial credit. Note that the categories falling between "unacceptable, wrong, off-task, or omitted" and "complete" represent increasing levels of a partially correct response.

IRT item parameter estimates were obtained from the BILOG/PARSCALE program with prior distributions imposed on item parameters (see Mislevy & Bock, 1982). Starting values for the item parameters were calculated in item analysis routines. The ability distribution over subjects was assumed fixed (normal [0,1]) and stable estimates of the IRT item parameters were obtained. Next, the parameter estimates from this initial solution were used as starting values for subsequent estimation in which the ability distribution over subjects was unconstrained and estimated concurrently with item parameter estimates.

Table 3-2
NAEP 1997 Arts Assessment
Score Levels for Polytomous Music Items

Score During the Rating Process	Score During the Scaling Process	2-Category Item Score	3-Category Item Score	4-Category Item Score	5-Category Item Score
5	4	†	†	†	Complete
4	3	†	†	Complete	Adequate
3	2	†	Complete	Essential	Uneven
2	1	Complete	Partial	Partial	Minimal
1	0	*	*	*	*

† Not applicable

* Unacceptable, wrong, off-task, or omitted

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The calibration process used student weights that were normalized so that the sum of weights equaled the sample size. Items that received special treatment in the scaling procedure are listed in Table 3-3. If items had empirical item response functions that were severely nonmonotonic, they were dropped from analysis. If polytomous items had sparse or nonmonotonic responses in one or more categories, adjacent score categories were collapsed (i.e., responses scored in adjacent categories were combined into a single category). If items were not independent from one another, they were combined together to create a clustered item where the total scores for each item were added together. One hundred three items were given special treatment. Thirteen items were dropped due to lack of fit and twenty-four were collapsed to improve item fit. Two items were dropped but were used in the *Creating* scale, which was not part of the IRT scale. In addition, item clustering is specified in Table 3-3.

Table 3-3
NAEP 1997 Arts Assessment
Music Items Receiving Special Treatment

NAEP ID	Block	Treatment
UC000A1	MC	Clustered with UC000B1
UC000B1	MC	Clustered with UC000A1
UCCL001	MC	Collapsed: (0,1,2,3) becomes (0,0,1,2)
UC000A2	MC	Clustered with UC000B2
UC000B2	MC	Clustered with UC000A2
UCCL002	MC	Collapsed: (0,1,2) becomes (0,1,1)
UC000A3	MC	Clustered with UC000B3
UC000B3	MC	Clustered with UC000A3
UCCL003	MC	Collapsed: (0,1,2,3,4) becomes (0,1,2,2,2)
UC00006	MC	Collapsed: (0,1,2) becomes (0,1,1)
UC000A8	MC	Clustered with UC000B8, UC000C8, UC000D8, UC000E8, UC000F8, and UC000G8
UC000B8	MC	Clustered with UC000A8, UC000C8, UC000D8, UC000E8, UC000F8, and UC000G8

(continued)

Table 3-3 (continued)
NAEP 1997 Arts Assessment
Music Items Receiving Special Treatment

NAEP ID	Block	Treatment
UCCL003	MC	Collapsed: (0,1,2,3,4) becomes (0,1,2,2,2)
UC00006	MC	Collapsed: (0,1,2) becomes (0,1,1)
UC000A8	MC	Clustered with UC000B8, UC000C8, UC000D8, UC000E8, UC000F8, and UC000G8
UC000B8	MC	Clustered with UC000A8, UC000C8, UC000D8, UC000E8, UC000F8, and UC000G8
UC000C8	MC	Clustered with UC000A8, UC000B8, UC000D8, UC000E8, UC000F8, and UC000G8
UC000D8	MC	Clustered with UC000A8, UC000B8, UC000C8, UC000E8, UC000F8, and UC000G8
UC000E8	MC	Clustered with UC000A8, UC000B8, UC000D8, UC000D8, UC000F8, and UC000G8
UC000F8	MC	Clustered with UC000A8, UC000B8, UC000C8, UC000D8, UC000E8, and UC000G8
UC000G8	MC	Clustered with UC000A8, UC000B8, UC000C8, UC000D8, UC000E8, and UC000F8
UCCL008	MC	Collapsed: (0,1,2,3,4,5,6,7) becomes (0,1,2,2,2,2,3,3)
UC00010	MC	Collapsed: (0,1,2,3) becomes (0,1,2,2)
UC00A14	MC	Clustered with UC00B14 and UC00C14
UC00B14	MC	Clustered with UC00A14 and UC00C14
UC00C14	MC	Clustered with UC00A14 and UC00B14
UCCL014	MC	Collapsed: (0,1,2,3,4,5) becomes (0,1,1,2,2,2)
UC00A15	MC	Clustered with UC00B15, UC00C15, and UC00D15
UC00B15	MC	Clustered with UC00A15, UC00C15, and UC00D15
UC00C15	MC	Clustered with UC00A15, UC00B15, and UC00D15
UC00D15	MC	Clustered with UC00A15, UC00B15, and UC00C15
UCCL015	MC	Collapsed: (0,1,2,3,4,5,6,7) becomes (0,0,0,1,1,2,3,4)
UD000D5	MD	Clustered with UD000E5
UD000E5	MD	Clustered with UD000D5
UDCL005	MD	Collapsed: (0,1,2,3) becomes (0,0,1,2)
UD000D6	MD	Clustered with UD000E6
UD000E6	MD	Clustered with UD000D6
UDCL006	MD	Collapsed: (0,1,2,3,4) becomes (0,0,0,1,2)
UD00007	MD	Collapsed: (0,1,2) becomes (0,1,1)
UD000A9	MD	Clustered with UD000B9
UD000B9	MD	Clustered with UD000A9
UDCL009	MD	Collapsed: (0,1,2,3) becomes (0,0,0,1)
UE000A1	ME	Clustered with UE000B1 and UE000C1
UE000B1	ME	Clustered with UE000A1 and UE000C1
UE000C1	ME	Clustered with UE000A1 and UE000B1
UECL001	ME	Collapsed: (0,1,2,3) becomes (0,1,1,1)
UE00004	ME	Collapsed: (0,1,2,3) becomes (0,1,1,2)
UE000A6	ME	Clustered with UE000B6
UE000B6	ME	Clustered with UE000A6
UE000A8	ME	Dropped due to lack of fit; clustered with UE000B8
UE000B8	ME	Dropped due to lack of fit; clustered with UE000A8

(continued)

Table 3-3 (continued)
NAEP 1997 Arts Assessment
Music Items Receiving Special Treatment

NAEP ID	Block	Treatment
UECL008	ME	Dropped and moved to <i>Creating Scale</i>
UE000C8	ME	Dropped and moved to <i>Creating Scale</i>
UE000A9	ME	Clustered with UE000B9 and UE000C9
UE000B9	ME	Clustered with UE000A9 and UE000C9
UE000C9	ME	Clustered with UE000A9 and UE000B9
UECL009	ME	Collapsed: (0,1,2,3,4,5,6) becomes (0,1,2,3,3,3,3)
UE00A11	ME	Clustered with UE00B11
UE00B11	ME	Clustered with UE00A11
UECL011	ME	Collapsed: (0,1,2,3) becomes (0,1,2,2)
UE00A12	ME	Clustered with UE00B12
UE00B12	ME	Clustered with UE00A12
UECL012	ME	Collapsed: (0,1,2,3) becomes (0,1,1,2)
UE00013	ME	Dropped due to lack of fit
UF000A1	MF	Clustered with UF000B1
UF000B1	MF	Clustered with UF000A1
UFCL0A1	MF	Dropped due to lack of fit
UF000C1	MF	Dropped due to lack of fit
UF000E1	MF	Dropped due to lack of fit
UF000G1	MF	Clustered with UF000H1
UF000H1	MF	Clustered with UF000G1
UFCL0G1	MF	Dropped due to lack of fit
UF000I1	MF	Dropped due to lack of fit
UF000K1	MF	Clustered with UF000L1
UF000L1	MF	Clustered with UF000K1
UFCL0K1	MF	Dropped due to lack of fit
UF000M1	MF	Clustered with UF000N1
UF000N1	MF	Clustered with UF000M1
UFCL0M1	MF	Dropped due to lack of fit
UF000O1	MF	Clustered with UF000P1
UF000P1	MF	Clustered with UF000O1
UFCL0O1	MF	Dropped due to lack of fit
UF000Q1	MF	Dropped due to lack of fit
UF000S1	MF	Clustered with UF000T1
UF000T1	MF	Clustered with UF000S1
UFCL0S1	MF	Dropped due to lack of fit
UF000A2	MF	Clustered with UF001B6 and UF000B2
UF000B2	MF	Clustered with UF001A6 and UF000A2
UFCL002	MF	Collapsed: (0,1,2,3) becomes (0,0,1,1)
UF00003	MF	Collapsed: (0,1,2,3) becomes (0,1,2,2)

(continued)

Table 3-3 (continued)
NAEP 1997 Arts Assessment
Music Items Receiving Special Treatment

NAEP ID	Block	Treatment
UF00005	MF	Collapsed: (0,1,2,3,4) becomes (0,1,2,2,3)
UF001A6	MF	Clustered with UF001B6 and UF001C6
UF001B6	MF	Clustered with UF001A6 and UF001C6
UF001C6	MF	Clustered with UF001A6 and UF001B6
UF002A6	MF	Clustered with UF002B6 and UF002C6
UFCL0A6	MF	Collapsed: (0,1,2,3,4,5,6,7) becomes (0,0,0,0,1,2,3,4)
UF002B6	MF	Clustered with UF002A6 and UF002C6
UF002C6	MF	Clustered with UF002A6 and UF002B6
UFCL0B6	MF	Collapsed: (0,1,2,3,4,5,6,7) becomes (0,0,1,2,3,4,5,6)
UF003A6	MF	Clustered with UF003B6 and UF003C6
UF003B6	MF	Clustered with UF003A6 and UF003C6
UF003C6	MF	Clustered with UF003A6 and UF003B6
UFCL0C6	MF	Collapsed: (0,1,2,3,4,5,6,7) becomes (0,0,1,2,3,4,5,6)
UF004A6	MF	Clustered with UF004B6 and UF004C6
UF004B6	MF	Clustered with UF004A6 and UF004C6-
UF004C6	MF	Clustered with UF004A6 and UF004B6
UFCL0D6	MF	Collapsed: (0,1,2,3,4,5,6,7) becomes (0,0,0,1,1,2,3,3)

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Unlike the Arts Assessment in Visual Arts and Theatre, which had univariate IRT scales, the Arts Assessment in Music had two subscales in the IRT model. These were combined to form a single weighted composite at the conditioning phase, which is discussed below (see Section 3.2.5).

3.2.3 Estimation of Group Score Distributions

For the assessment sample, group score distributions were estimated for the two Music proficiency scales, using the multivariate program CGROUP. As with scaling, student weights were used at this stage of the analysis. Instead of using the actual background variables for the estimation (as had been done prior to the 1990 assessment), principal components of the background variables were used. The principal components employed accounted for 90 percent of the variance of the original background variables (as coded in Appendix D) selected for estimation. Principal components were employed to remedy problems of extreme collinearity among some of the original variables.

Research based on data from the 1990 Trial State Assessment suggests that results obtained using the 90 percent subset of components will differ only slightly from those obtained using the full set (Mazzeo, Johnson, Bowker, & Fong, 1992). Table 3-4 contains a list of the number of principal components, as well as the proportion of scale score variance accounted for by the estimation model for grade 8. A list of the variables defining group membership can be found in Appendix G.

Table 3-4
NAEP 1997 Arts Assessment in Music
Proportion of Scale Score Variance
Accounted for by the Estimation Model

Number of Contrasts ¹	Number of Principal Components ¹	Proportion of Scale Score Variance Accounted for by the Estimation Model
459	145	.61

¹ excluding the constant term

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The CGROUP program estimates distributions of scale scores by combining information from item responses (i.e., likelihood information) and background variables (i.e., prior information) of individuals.

3.2.4 TRANSFORMATION OF THE MUSIC CALIBRATION SCALES FOR REPORTING

Since the NAEP 1997 Arts Assessment in Music was developed and scaled using within-grade procedures, and since there was no prior Arts Assessment in Music with a comparable framework to which it was being linked, a new reporting metric was adopted. The results are reported on 0–300 scale. The same convention was adopted for the Theatre and Visual Arts assessments. As is shown in Table 3-5, the mean of the two Music subscales was set at 150 for each grade, and the standard deviation at 35. After transformation a composite Music scale was created as a weighted sum of the two subscales. For each imputed value, scale 1 was weighted by .6 and scale 2 by .4 and then the products were summed together.

Table 3-5
NAEP 1997 Arts Assessment Means and Standard Deviations
of All Five Plausible Values for the Music Responding Scales

Scale	Mean	S. D.
Scale 1	150.00	35.00
Scale 2	150.00	35.00
Composite	150.00	33.72

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

If the achievement distribution were normal, we would expect this range (0–300) to cover about 99.998 percent of the distribution. Note that any transformed scale scores below 0 were censored to values of 0, and any transformed scale scores greater than 300 were censored to values of 300.

In order to put scores onto a scale with a mean of 150 and a standard deviation of 35, a linear transformation was applied. This means that each score was multiplied by one constant (A) and then added to another constant (B), as illustrated by the formula

$$\theta_{target} = A \times \theta_{calibrated} + B$$

where A and B are linear transformation constants. The values of A and B for each scale are given in Table 3-6.

Table 3-6
*NAEP 1997 Arts Assessment Transformation Constants
for the Music Responding Scales*

	A	B
Scale 1	35.40	150.00
Scale 2	33.94	150.76

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

3.2.5 Computation of the Music Composite Scale

The composite scale for Music was created by summing the Scale 1 score and the Scale 2 score, weighting the Scale 1 score by 0.6 and the Scale 2 score by 0.4. The weighting reflects the relative emphasis of each scale in the assessment. As seen in Table 3-5, the mean of the composite scale scores is 150, the same as the means of Scales 1 and 2. The standard deviation of the composite scale scores is less than 35, however, due to its calculation as an average of two correlated scales with standard deviation of 35. Scores on this transformed scale are used as a basis for making mean and percentile comparisons between subgroups for the 1997 Arts Assessment in Music. One should note, however, that comparisons between assessment disciplines (e.g., Theatre versus Music) do not have meaning.

3.3 CALCULATING THE MEAN PERCENT-CORRECT SCALES FOR MUSIC CREATING/PERFORMING ITEMS

Table 3-7 gives information similar to Table 3-1, but in regard to the *Creating* and *Performing* items, that is, items that require that students be involved in some sort of creating or performing activity (e.g., improvising or playing a piece on an instrument). These items were distinguished from *Responding* items. *Responding* items were reported using an IRT scale. *Performing* items were reported using a percent of total possible score for each administered session type (see Section 3.3.1). Table 3-7 gives

item statistics for the blocks created exclusively with *Creating* and *Performing* items. This table shows that mean item scores, average r-Polyserials, and reliabilities of the block varied. The proportion of students attempting the last item in the block is not meaningful because the “items” contributing to each block can be scores for different aspects of the same performance.

Table 3-7
*NAEP 1997 Arts Assessment Descriptive Statistics
for the Creating and Performing Items by Block for the Music Sample*

Statistic	MG	MH	MI
Number of Items	5	8	7
Unweighted Sample Size	659	628	668
Weighted Mean Item Score Averaged Over Block	.31	.36	.44
Weighted Alpha Reliability	.77	.87	.76
Weighted Average r-Polyserial	.81	.85	.70

NOTE: Some of the Arts Assessment items were scored several times to evaluate different aspects of responses. The items specified in this table include these separate scorings of responses as separate items.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Tables D-1 and D-2 in Appendix D list the interrater reliability information for constructed-response *Creating* and *Performing* items in blocks MG-MI and the two *Creating* items in block ME (one of these is a cluster item). All but two of the items in the *Creating* and *Performing* scales were polytomously-scored constructed-response items. The reliability, indicated by proportion exact agreement, for these polytomously scored items ranged from .60 to .91 and averaged .77. Though in a lower range than comparable to reliabilities for past Reading, Mathematics, and Writing items, considering the difficulty of rating music performances we consider this range of reliability acceptable. Cohen’s Kappa reliabilities (most appropriate for dichotomized items) and intraclass correlations (appropriate for polytomous items) correct reliability estimates for chance matches. The intraclass correlations for the polytomous items were somewhat higher than the percent exact agreement, ranging from .66 to .95 with an average of .86.

3.3.1 The Percent-Correct Scales for *Creating* and *Performing* Items

As mentioned above, the *Creating* and *Performing* items were not scaled using IRT methods. These items were used to form descriptive scales comprised of percent-correct scores of *Performing* and *Creating* items. A session is a combination of Music blocks administered to a subsample of students. Sessions were spiraled so that equivalent populations responded to each item. Table 3-8 shows the session structure of the Arts Assessment in Music. During each session, students were given a book of *Responding* blocks (also called A/B blocks) and some combination of *Creating* or *Performing* blocks. Also, a small subsample of students participating in the Arts Assessment in Music were identified as being involved in a regular music activity such as singing in the school chorus or taking music lessons. These students were given a special *Performing* block, as indicated in column 4 of Table 3-8, and

contributed to the “self-selected sample.” Of the non-*Responding* items, some were used to form a *Creating* scale and others were used to form a *Performing* scale. Items in the *Creating* scale involved students in a creative task such as writing an ending to a rhythmic phrase, creating a rhythmic embellishment based on a familiar tune, developing an original melody, and participating in improvisation activities. On the other hand, items in the *Performing* scale called on students to perform a musical piece (e.g., play a familiar tune by ear on the MIDI keyboard or sing the song “America” with audiotape accompaniment).

Table 3-8
NAEP 1997 Arts Assessment
Session/Block Structure for Music

Session	<i>Responding</i> Blocks		<i>Creating/ Performing</i> Blocks	<i>Creating/Performing</i> Blocks for Students in a Music Activity ¹
1	MC	MC	MG	MJ
2	ME	MC	MH	MK
3	MC	MF	MI	MJ
4	MD	ME	MI	MK
5	MF	MD	MH	MJ
6	ME	MF	MG	MK
7	MD	MC	MG	MJ
8	MD	MC	MH	MK
9	MF	ME	MH	MJ
10	MF	ME	MI	MK

¹ Only self-selected students from the session received these blocks of items. The items in these blocks were not included in the *Creating* or *Performing* scales, but results for these items were reported at an individual item level.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Tables 3-9 and 3-10 show the items that formed each scale. Note that items in blocks MJ and MK, which were administered to students who were also involved in a music activity, were not used to form a *Creating* or *Performing* scale but were only reported on at the individual item level.

Table 3-9
NAEP 1997 Arts Assessment
Items in the Music Creating Scale

Block	Item
ME	UECL008
ME	UE000C8
MG	UG000A1
MG	UG000B1
MG	UG000C1
MH	UH000C1
MI	UI000A1
MI	UI000C1
MI	UI000D1
MI	UI000E1
MI	UI000A3
MI	UI000B1
MI	UI000C3

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table 3-10
NAEP 1997 Arts Assessment
Items in the Music Performing Scale

Block	Item
MG	UG000A2
MG	UG000B2
MH	UH000A1
MH	UH000B1
MH	UH000A2
MH	UH000B2
MH	UH000C2
MH	UH000D2
MH	UH000E2

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The *Performing* or *Creating* percent-correct score for a student is the total score of the student for the session divided by the maximum possible score for the session. It should be noted that these scales are descriptive, as different students received different items of varying difficulty. However, since the sessions were spiraled, all relevant subgroups were equally exposed to all session types. The percent-correct scale can be used for making meaningful comparisons among subgroups of students in this assessment.

Section 4

DATA ANALYSIS FOR THE 1997 ARTS ASSESSMENT IN THEATRE⁶

Frank Jenkins, Spencer S. Swinton, Laura J. Jerry, Edward Kulick, and Venus Leung
Educational Testing Service

4.1 INTRODUCTION

This section describes the analyses of the cognitive and background item responses in the NAEP 1997 Arts Assessment in Theatre at grade 8. This section focuses on the methods and procedures used to estimate distributions of IRT-based scale scores⁷ and percent-correct summaries for subgroups of students. The wide array of topics include, for example, the scoring of constructed-response items, classical item statistics, mean item scores, item response theory (IRT) analysis of the Theatre scale, and estimation of subgroup means through marginal maximum likelihood estimation of group score distributions. The statistical bases of the IRT and plausible values methodology used in NAEP and described in this section are given in the NAEP 1996 Technical Report (Allen, Carlson, & Zelenak, 1999) and the NAEP 1998 Technical Report (Allen, Donoghue, & Schoeps, 2001). These analyses serve as a basis for the results presented in the Theatre sections of the NAEP Arts Report Card (Persky, Sandene, & Askew, 1998).

There were two major steps in the analysis of the NAEP 1997 Theatre data, each of which is described in a separate section:

- creating the item response theory (IRT) scales for the *Responding* items (see Section 4.2) and
- calculating the mean percent-correct scales for the *Creating/Performing* items (see Section 4.3).

4.2 CREATING THE ITEM RESPONSE THEORY (IRT) SCALE FOR THEATRE *RESPONDING* ITEMS

4.2.1 Item Analysis

This section contains a description of the item analysis performed using the sample data from selected schools. The analysis examines items within blocks. In preparation for this step, constructed-response items were scored by persons trained as raters, and derived background variables were calculated. Item statistics such as mean item score, average block score, item to total score correlations and percent responding in each item category were calculated.

⁶ Spencer S. Swinton was primarily responsible for the planning, specification, and coordination of the Theatre analyses. Computing activities for all Theatre scaling and data analyses were directed by Edward Kulick and completed by Laura J. Jerry and Venus Leung. Others contributing to the analysis of Theatre data were David S. Freund, Bruce A. Kaplan, and Katharine E. Pashley.

⁷ NAEP scale scores are based on item response theory (IRT).

In NAEP analyses (see Section 2) a distinction is made between missing responses at the end of each block (not reached) and missing responses prior to the last completed response (omitted). Not-reached items are those occurring after the last item a student answers in a block. Items that were not reached were treated as if they had not been presented to the student, and were therefore not counted as wrong. An omission occurs when a student skips an item but responds to some item(s) after it. Omitted items were scored as incorrect. The proportion of students attempting the last item of a block is often used as an index of the degree of speededness of the block of items (i.e., the smaller the proportion of students reaching the end of the block, the more speeded the block).

Table 4-1 shows block statistics for *Responding* items, that is, standard multiple-choice or constructed-response items presented in a paper and pencil format, which were analyzed with IRT methods. In contrast, *Creating/Performing* items are those which require participation in a theatrical performance task (they will be described in Section 4.3).

Table 4-1
NAEP 1997 Arts Assessment Descriptive Statistics
for the IRT-Scaled Responding Items by Block for the Theatre Sample

Statistic	TC	TD	TE	TF
Number of Items	11	9	7	14
Number of Scaled Items ¹	6	4	6	11
Number of Scaled Constructed-Response Items ¹	5	2	6	8
Number of Scaled Multiple-Choice Items ¹	1	2	0	3
Unweighted Sample Size ¹	633	674	699	750
Weighted Mean Item Score Averaged Over Block ¹	.56	.56	.58	.48
Weighted Alpha Reliability ¹	.61	.45	.66	.71
Weighted Average r-Polyserial ¹	.70	.84	.71	.62
Weighted Proportion of Students Attempting Last Item	.99	.89	.90	.98

¹ The *Creating/Performing* items from blocks TC & TE are deleted. Some of the Arts Assessment items were scored several times to evaluate different aspects of responses. The items specified in this table include these separate scorings of responses as separate items.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The table shows the number of scaled items, number of scaled constructed-response items, number of scaled multiple-choice items, sample size, mean item score, alpha reliability, mean item to total score correlation, and the proportion of students attempting the last item in the block for each block. These values were calculated within block only for those items used in the scaling process. Student sample weights were used to calculate statistics, except for the sample size. The results indicated that the blocks differ in number of items, average difficulty, reliability, and percent reaching the last item. TF is the easiest block (mean item score = .78) with low speededness (98% reaching the last item) with the highest reliability (alpha = .71). In contrast, the other blocks have about the same difficulty (mean item score = .56 or .58), with the lowest reliability of alpha = .45. The most speeded block was TD (89% reaching the last item). Approximately 25 percent of the responses to constructed-response items were

scored twice to assess interrater reliability. Tables D-3 and D-4 in Appendix D list the reliability information. Dichotomous constructed-response items were scored on a scale of 1 to 2, while polytomous constructed-response items were scored on a scale of 1 to 3, 1 to 4, or 1 to 5 to reflect degrees of knowledge. In IRT parameter estimation, this scale is shifted down by 1 (see Table 4-2). Table D-4 in Appendix D indicates that for the *Responding* items, proportion exact agreement for constructed-response items was in a range comparable to other assessments with values from .79 to .96 (averaging .86). The intraclass correlations (appropriate for polytomous items) are reliability estimates that correct for chance agreement. These also fell into an acceptable range with values from .73 to .95 and an average of .83.

4.2.2 Estimation of Item Parameters

For each arts discipline (Music, Theatre, and Visual Arts), a separate IRT scale was constructed using the NAEP BILOG/PARSCALE program (Mislevy & Bock, 1982; Muraki & Bock, 1991). The program uses marginal estimation procedures to estimate the parameters of the one-, two-, and three-parameter logistic models, and the generalized partial credit model (Muraki, 1992). Dichotomous multiple-choice items were analyzed using a three-parameter IRT model; dichotomized constructed-response items employed a two-parameter IRT model. Three- and four-category items were polytomously scored and were analyzed with a generalized partial credit model (Muraki, 1992).

Creating/Performing items were written to different specifications than *Responding* items, and evidenced dependencies that would violate the local independence assumption of IRT. As a result, they were not included in the IRT scale, but were used to form a separate scale based on a percent-correct score (see Section 4.3).

As stated earlier, item responses that were missing prior to the last completed item in a block were considered omitted and scored as wrong. Also, items that were not reached were treated as if they were not presented to the students (and not counted as wrong). For IRT scaling omitted multiple-choice items were treated as fractionally correct ($1 \div [\text{number of alternatives}]$). Responses to constructed-response items that were off-task were treated as omitted and assigned to the lowest category. For Theatre, constructed-response items had two, three, or four score categories of increasing credit. Table 4-2 lists the score level definitions for constructed-response items which were scored with partial credit. Note that the categories falling between “unacceptable, wrong, off-task, or omitted” and “complete” represent increasing levels of a partially correct response.

Table 4-2
NAEP 1997 Arts Assessment
Score Levels for Polytomous Theatre Items

Score During the Rating Process	Score During the Scaling Process	2-Category Item Score	3-Category Item Score	4-Category Item Score
4	3	†	†	Complete
3	2	†	Complete	Essential
2	1	Complete	Partial	Partial
1	0	*	*	*

† Not applicable

* Unacceptable, wrong, off-task, or omitted

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

IRT item parameter estimates were obtained from the BILOG/PARSCALE program with prior distributions imposed on item parameters (see Mislevy & Bock, 1982). Starting values for the item parameters were calculated in item analysis routines. The ability distribution over subjects was assumed fixed (normal [0,1]) and stable estimates of the IRT item parameters were obtained. Next, the parameter estimates from this initial solution were used as starting values for subsequent estimation in which the ability distribution over subjects was unconstrained and estimated concurrently with item parameter estimates.

The calibration process used student weights that were normalized so that the sum of weights equaled the sample size. Items that received special treatment in the scaling procedure are listed in Table 4-3, along with the reason for special treatment. If items had empirical item response functions that were severely nonmonotonic, they were dropped from analysis. If polytomous items had sparse or nonmonotonic responses in one or more categories, adjacent score categories were collapsed (i.e., responses scored in adjacent categories were combined into a single category). Twenty-two items were given special treatment, eight were collapsed to improve fit, and two were dropped due to lack of fit. Nine additional items were taken out of the IRT scale and were used in the *Creating/Performing* scale (see Section 4.3.1). A list of the variables defining group membership can be found in Appendix G.

Table 4-3
NAEP 1997 Arts Assessment
Theatre Items Receiving Special Treatment

NAEP ID	Block	Treatment
HC00002	TC	Collapsed: (0,1,2) becomes (0,0,1)
HC00003	TC	Dropped due to lack of fit
HC000A5	TC	Non-cognitive item; not scaled
HC000B5	TC	Dropped and moved to <i>Creating/Performing</i> scale
HC000C5	TC	Dropped and moved to <i>Creating/Performing</i> scale
HC00006	TC	Collapsed: (0,1,2,3) becomes (0,1,1,2)
HC00007	TC	Dropped and moved to <i>Creating/Performing</i> scale
HD00003	TD	Dropped due to lack of fit
HD00005	TD	Dropped and moved to <i>Creating/Performing</i> scale
HD00006	TD	Dropped and moved to <i>Creating/Performing</i> scale
HD00007	TD	Dropped and moved to <i>Creating/Performing</i> scale
HD00008	TD	Dropped and moved to <i>Creating/Performing</i> scale
HD00009	TD	Collapsed: (0,1,2) becomes (0,0,1)
HE00001	TE	Collapsed: (0,1,2) becomes (0,0,1)
HE00002	TE	Collapsed: (0,1,2) becomes (0,0,1)
HE00004	TE	Dropped and moved to <i>Creating/Performing</i> scale
HE00006	TE	Collapsed: (0,1,2) becomes (0,0,1)
HF000A1	TF	Non-cognitive item; not scaled
HF00004	TF	Collapsed: (0,1,2) becomes (0,0,1)
HF00006	TF	Dropped and moved to <i>Creating/Performing</i> scale
HF00007	TF	Collapsed: (0,1,2,3) becomes (0,1,2,2)
HF00A12	TF	Non-cognitive item; not scaled

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

4.2.3 Estimation of Group Score Distributions

For the assessment sample, univariate group score distributions were estimated for a single Theatre proficiency scale, using the program BGROUP. As with scaling, student weights were used at this stage of the analysis. Instead of using the actual background variables for the estimation (as had been done prior to the 1990 assessment), principal components of the background variables were used. The principal components employed accounted for 90 percent of the variance of the original background variables (as coded in Appendix D) selected for the estimates. Principal components were employed to remedy problems of extreme collinearity among some of the original variables.

Research based on data from the 1990 Trial State Assessment suggests that results obtained using the 90 percent subset of components will differ only slightly from those obtained using the full set (Mazzeo, Johnson, Bowker, & Fong, 1992). Table 4-4 contains a list of the number of principal components, as well as the proportion of scale score variance accounted for by the estimation model for grade 8. A list of the variables defining group membership can be found in Appendix D.

Table 4-4
1997 Arts Assessment in Theatre
Proportion of Scale Score Variance
Accounted for by the Estimation Model

Number of Contrasts ¹	Number of Principal Components ¹	Proportion of Scale Score Variance Accounted for by the Estimation Model
459	145	.61

¹ Excluding the constant term.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The BGROUP program estimates distributions of scores by combining information from item responses (i.e., likelihood information) and background variables (i.e., prior information) of individuals. For each individual, five plausible values are randomly drawn from their estimated posterior distribution.

4.2.4 Transformation of the Theatre Calibration Scale for Reporting

Since the 1997 Arts Assessment in Theatre was developed and scaled using within-grade procedures, and since there was no prior Arts Assessment in Theatre with a comparable framework to which it was being linked, a new reporting metric was adopted. The results are reported on 0–300 scale. The same convention was adopted for the Arts Assessments in Music and Visual Arts. As is shown in Table 4-5, the mean of the Theatre scale was set at 150 for each grade, and the standard deviation at 35.

If the achievement distribution was normal, we would expect this range (0–300) to cover about 99.998 percent of the distribution. Note that any transformed scale scores below 0 were censored to values of 0, and any transformed scale scores greater than 300 were censored to values of 300.

Table 4-5
NAEP 1997 Arts Assessment
Means and Standard Deviations
of All Five Plausible Values
for the Theatre Responding Scale

Mean	S. D.
150	35

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Scores on this transformed scale are used as a basis for making mean and percentile comparisons between subgroups for the NAEP 1997 Arts Assessment in Theatre. One should note, however, that comparisons between one assessment discipline (e.g., Theatre) and another (e.g., Music) do not have meaning.

In order to put scores onto a scale with a mean of 150 and a standard deviation of 35, a linear transformation is applied to the scores. This means that each score was multiplied by one constant (A) and then added to another constant (B), as illustrated by the formula

$$\theta_{target} = A \times \theta_{calibrated} + B$$

where A and B are linear transformation constants. The values of A and B for the Theatre sample are given in Table 4-6.

Table 4-6
*NAEP 1997 Arts Assessment
Transformation Constants
for the Theatre Responding Scale*

A	B
34.86	148.88

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

4.3 CALCULATING THE MEAN PERCENT-CORRECT SCALE FOR THEATRE CREATING/PERFORMING ITEMS

Table 4-7 gives information similar to Table 4-1, but for the *Creating/Performing* items. These items require that students be involved in a *Creating/Performing* activity, such as acting in a scene. These items were distinguished from *Responding* items. *Responding* items were reported using an IRT scale. *Creating/Performing* items were reported using a percent of total possible score for each administered session type (see Section 4.3.1). Table 4-7 gives item statistics for the blocks made up exclusively of *Creating/Performing* items. This table shows that mean item scores, average r-Polyserials, and reliabilities of the blocks varied quite a bit.

Tables D-3 and D-4 in Appendix D list the interrater reliability information for constructed-response *Creating/Performing* items in blocks TG–TI and eight of the nine *Creating/Performing* items in blocks TC–TF, as well as for the constructed-response *Responding* items. (There was one multiple-choice item in block TD that was included in this scale.) The reliability, indicated by proportion exact agreement was almost always above .80 for all but two of the *Creating/Performing* items and averaged .88. The range of values was from .77 to .97. This is an acceptable range, comparable to reliabilities for past assessments such as Reading, Mathematics, and Writing. All of the constructed-response *Creating/Performing* items had more than two categories of responses; therefore, intraclass correlations

were also examined. These, too, were comparable to other assessments, ranging from .70 to .96 with an average of .86.

Table 4-7
*NAEP 1997 Arts Assessment Descriptive Statistics
for the Creating/Performing Items by Block for the Theatre Sample*

Statistic	TG	TH	TI
Number of Items in Scale	6	5	6
Unweighted Sample Size	431	385	390
Weighted Mean Item Score Averaged Over Block	.62	.48	.42
Weighted Alpha Reliability	.62	.89	.71
Weighted Average r-Polyserial	.72	.94	.83

NOTE: Some of the Arts Assessment items were scored several times to evaluate different aspects of responses. The items specified in this table include these separate scorings of responses as separate items.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

4.3.1 The Percent-Correct Scale for *Creating/Performing* Items

As mentioned above, the *Creating/Performing* items were not scaled using IRT methods. These items were used to form a descriptive scale comprised of percent-correct scores of *Creating/Performing* items for each session type. A session is a combination of Theatre blocks administered to a subsample of students. Sessions were spiraled so that equivalent populations responded to each item. Table 4-8 shows the session structure of the NAEP 1997 Arts Assessment in Theatre.

Table 4-8
*NAEP 1997 Arts Assessment
Session/Block Structure for Theatre*

Session	<i>Responding Blocks</i>		<i>Creating/ Performing Blocks</i>	Number of <i>Creating/Performing Items</i>
1	TC	TD	TG	13
2	TE	TC	TH	9
3	TC	TF	TI	10
4	TD	TE	TI	11
5	TF	TD	TH	10
6	TE	TF	TG	8
7	TD	TC	TG	13
8	TD	TC	TH	12
9	TF	TE	TH	7
10	TF	TE	TI	8

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Students in each session received a booklet of *Responding* blocks (also called A/B blocks), and students in some sessions also received a *Creating/Performing* block. Note that all of the *Responding* blocks contain some items that are considered *Creating/Performing* items. These items are not included in the IRT scale but *are* included in the percent-correct *Creating/Performing* scale. The fourth column of Table 4-8 shows the number of *Creating/Performing* items in a session, which is the sum of the *Creating/Performing* items from the *Responding* blocks (TC, TD, TE, TF) and the *Creating/Performing* items from the *Creating/Performing* blocks (TG, TH, TI). The *Creating/Performing* percent-correct score for a student is the total score of the student for the session divided by the maximum possible score for the session multiplied by 100. For example, if the maximum possible score on these items is 11, if a student got a total score of 5 on the session his or her percent-correct score would be $(5 \div 11) \times 100 = 45$ percent. It should be noted that these scales are descriptive, as different students received different items of varying difficulty. However, since the sessions were spiraled, all relevant subgroups were equally exposed to all session types. The percent-correct scale can be used for making meaningful comparisons among subgroups of students in this assessment. Some items in blocks TG–TI, which were administered to students who were also involved in a theatrical activity, were not used to form the *Creating/Performing* scale but were only reported on at the individual item level.

Table 4-9 lists the items that were included in the Theatre *Creating/Performing* scale.

Table 4-9
*NAEP 1997 Arts Assessment Items in the
Theatre Creating/Performing Scale*

Block	Item
TC	HC000B5
TC	HC000C5
TC	HC00007
TD	HD00005
TD	HD00006
TD	HD00007
TD	HD00008
TE	HE00004
TF	HF00006
TG	HG000A1
TG	HG000B1
TG	HG000C1
TG	HG000D1
TG	HG000E1
TG	HG000F1
TH	HH000A1
TH	HH000B1
TH	HH000C1
TH	HH000D1

(continued)

Table 4-9 (continued)
*NAEP 1997 Arts Assessment Items in the
Theatre Creating/Performing Scale*

Block	Item
TH	HH000F1
TI	HI000A1
TI	HI000B1
TI	HI000C1
TI	HI000D1
TI	HI000E1
TI	HI000F1

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Section 5

DATA ANALYSIS FOR THE 1997 ARTS ASSESSMENT IN VISUAL ARTS⁸

Frank Jenkins, Spencer S. Swinton, Laura J. Jerry, Edward Kulick, and Xiaohui Wang
Educational Testing Service

5.1 INTRODUCTION

This section describes the analyses of the Visual Arts cognitive and background item responses in the NAEP 1997 Arts Assessment at grade 8. This section focuses on the methods and procedures used to estimate distributions of IRT-based scale scores⁹ and percent-correct summaries for subgroups of students. The wide array of topics include, for example, the scoring of constructed-response items, classical item statistics, mean percent scales, item response theory (IRT) analysis of the Visual Arts scale, and estimation of subgroup means through marginal maximum likelihood estimation of group score distributions. The statistical bases of the IRT and plausible values methodology used in NAEP and described in this section are given in the *NAEP 1996 Technical Report* (Allen, Carlson, & Zelenak, 1999) and the *NAEP 1998 Technical Report* (Allen, Donoghue, & Schoeps, 2001). These analyses serve as a basis for the results presented in the Visual Arts sections of the *NAEP Arts Report Card* (Persky, Sandene, & Askew, 1998).

There were two major steps in the analysis of the NAEP 1997 Visual Arts data, each of which is described in a separate section:

- creating the item response theory (IRT) scales for the *Responding* items (Section 5.2) and
- calculating the mean percent-correct scales for the *Creating* items (Section 5.3).

5.2 CREATING THE ITEM RESPONSE THEORY (IRT) SCALE FOR VISUAL ARTS *RESPONDING* ITEMS

5.2.1 Item Analysis

This section contains a description of the item analysis performed using nationally representative sample data. The analysis examines items within blocks. In preparation for this step, constructed-response items were scored by persons trained as raters, and derived background variables were calculated. Item statistics such as mean item score, average block score, item to total score correlations and percent responding in each item category were calculated.

⁸ Spencer S. Swinton was primarily responsible for the planning, specification, and coordination of the Visual Arts analyses. Computing activities for all Visual Arts scaling and data analyses were completed by Laura J. Jerry and Xiaohui Wang. Others contributing to the analysis of Visual Arts data were David S. Freund, Bruce A. Kaplan, Edward Kulick, and Katharine E. Pashley.

⁹ Scale scores based on item response theory (IRT)

In NAEP analyses (see Section 2) a distinction is made between missing responses at the end of each block (not-reached) and missing responses prior to the last completed response (omitted). Not-reached items are those occurring after the last item a student answers in a block. Items that were not reached were treated as if they had not been presented to the examinee, and were therefore not counted as wrong. An omit occurs when a student skips an item but responds to some item(s) after it. Omitted items are scored as incorrect. The proportion of students attempting the last item of a block is often used as an index of the degree of speededness of the block of items, i.e., the smaller the proportion of students reaching the end of the block, the more speeded the block.

Table 5-1 shows block statistics for *Responding* items, that is, standard multiple-choice or constructed-response items presented in a paper and pencil format, which were analyzed with IRT methods. In contrast, *Creating* items are those which require participation in some sort of creative task (They will be described in Section 5.3). The table shows the number of scaled items, number of scaled constructed-response items, number of scaled multiple-choice items, sample size, mean item score, alpha reliability, mean item to total score correlation, and the proportion of students attempting the last item in the block for each block. These values were calculated within block only for those items used in the scaling process. Student sample weights were used to calculate statistics, except for the sample size. The results indicated that the blocks differ in number of items, average difficulty, reliability, and percent reaching the last item. VF is the easiest block (mean item score = .51) with zero speededness (100% reaching the last item), but with the lowest reliability (alpha = .43). In contrast, block VC is the most difficult (mean item score = .42) with one of the higher reliabilities (alpha = .59) and is one of the most speeded blocks (82% reaching the last item).

Table 5-1
NAEP 1997 Arts Assessment Descriptive Statistics
for the IRT-Scaled Responding Items by Block for the Visual Arts Sample

Statistic	VC	VD	VE	VF
Number of Items Before Clustering	9	13	10	12
Number of Items After Clustering	8	13	10	12
Number of Scaled Items	5	12	9	6
Number of Scaled Constructed-Response Items	4	8	6	4
Number of Scaled Multiple-choice Items	1	4	3	2
Unweighted Sample Size	983	977	957	941
Weighted Mean Item Score Averaged Over Block	.42	.47	.48	.51
Weighted Alpha Reliability	.59	.64	.51	.43
Weighted Average r-Polyserial	.67	.51	.54	.60
Weighted Proportion of Students Attempting Last Item	.82	.90	.81	1.00

NOTE: The *Creating* items from blocks VC–VF are not described in this table. Some of the Arts Assessment items were scored several times to evaluate different aspects of responses. The items specified in this table include these separate scorings of responses as separate items.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Approximately 25 percent of the responses to constructed-response items were scored twice to assess interrater reliability. Appendix D lists the reliability information. Dichotomous constructed-response items were scored on a scale of 1 to 2, while polytomous constructed-response items were scored on a scale of 1 to 3, 1 to 4, or 1 to 5 to reflect degrees of knowledge. (In IRT parameter estimation, this scale is shifted down by 1 so that items have scores 0 to 2, 0 to 3, and 0 to 4, respectively.) The reliability for dichotomized responses was in an acceptable range comparable to other NAEP assessments, with proportion exact agreement ranging from .87 to .97 (averaging .91). The Cohen's Kappa statistic, which adjusts reliability estimates of dichotomous responses for chance agreement, ranged from .71 to .94 (averaging .80), as shown in Table D-5 in Appendix D. Table D-6 indicates that for the *Responding* polytomous items, proportion exact agreement was in an acceptable range of .80 to .94 (averaging .86). The intraclass correlation (appropriate for polytomous items) was also in an acceptable range of .75 to .97 (averaging .87).

5.2.2 Estimation of Item Parameters

For each arts discipline, a separate IRT scale was constructed using the NAEP BILOG/PARSCALE program (Mislevy & Bock, 1982; Muraki & Bock, 1991). The program uses marginal estimation procedures to estimate the parameters of the one-, two-, and three-parameter logistic models, and the generalized partial credit model (Muraki, 1992). For dichotomous multiple-choice items, a three-parameter IRT model was used, and for dichotomous constructed-response items, a two-parameter IRT model was used. Three- and four-category items were polytomously scored and were analyzed with a generalized partial credit model (Muraki, 1992).

Creating items were written to different specifications than *Responding* items, and evidenced dependencies that violate the local independence assumption of IRT. As a result, they were not included in the IRT scale, but were used to form a separate scale based on a percent-correct score (see Section 5.3).

As stated earlier, item responses that were missing prior to the last completed item in a block were considered omitted and scored as wrong. Also, items that were not reached were treated as if they were not presented to the examinees (and not counted as wrong). For IRT scaling, omitted multiple-choice items were treated as fractionally correct ($1 \div [\text{number of alternatives}]$). Responses to constructed-response items that were off-task were treated as omitted and assigned to the lowest category. For Visual Arts, constructed-response items had two, three, or four score categories. Table 5-2 lists the score level definitions for constructed-response items which were scored with partial credit. Note that the categories falling between “unacceptable” and “complete” represent increasing levels of a partially correct response.

Table 5-2
NAEP 1997 Arts Assessment
Score Levels for Polytomous Visual Arts Items

Score During the Rating Process	Score During the Scaling Process	2-Category Item Score	3-Category Item Score	4-Category Item Score
4	3	†	†	Complete
3	2	†	Complete	Essential
2	1	Complete	Partial	Partial
1	0	*	*	*

† Not applicable

* Unacceptable, wrong, off-task, or omitted

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

IRT item parameter estimates were obtained from the BILOG/PARSCALE program with prior distributions imposed on item parameters (see Mislevy & Bock, 1982). Starting values for the item parameters were calculated in item analysis routines. The ability distribution over subjects was assumed fixed (normal [0,1]), and stable estimates of the IRT item parameters were obtained. Next, the parameter estimates from this initial solution were used as starting values for subsequent estimation in which the ability distribution over subjects was unconstrained and estimated concurrently with item parameter estimates.

The calibration process used student weights that were normalized so that the sum of weights equaled the sample size. Items that received special treatment in the scaling procedure are listed in Table 5-3, along with the reason for special treatment. Items with severely nonmonotonic empirical item response functions were dropped from analysis. For any polytomous items that had sparse or nonmonotonic responses in one or more categories, adjacent score categories were collapsed (i.e., adjacent responses scored in categories were combined into a single category). Seventeen of the Visual Arts items were given special treatment. Two items were dropped due to lack of fit, and four were collapsed improve item fit. Also noted in Table 5-3, eight items were dropped but were used in the *Creating* scale, which was not part of the IRT scale.

Table 5-3
NAEP 1997 Arts Assessment
Visual Arts Items Receiving Special Treatment

NAEP ID	Block	Treatment
VC00001	VC	Dropped due to lack of fit
VC000A4	VC	Clustered with VC000B4
VC000B4	VC	Clustered with VC000A4
VC000A6	VC	Dropped and moved to <i>Creating</i> Scale
VC000B6	VC	Dropped and moved to <i>Creating</i> Scale

(continued)

Table 5-3 (continued)
NAEP 1997 Arts Assessment
Visual Arts Items Receiving Special Treatment

NAEP ID	Block	Treatment
VD00004	VD	Dropped due to lack of fit
VD00008	VD	Collapsed: (0,1,2,3) becomes (0,1,1,2)
VE00002	VE	Collapsed: (0,1,2,3) becomes (0,0,1,2)
VE00004	VE	Collapsed: (0,1,2) becomes (0,1,1)
VE00009	VE	Dropped and moved to <i>Creating</i> Scale
VF000A6	VF	Collapsed: (0,1,2,3) becomes (0,0,1,2)
VF000B6	VF	Dropped and moved to <i>Creating</i> Scale
VF000C6	VF	Dropped and moved to <i>Creating</i> Scale
VF000B7	VF	Dropped and moved to <i>Creating</i> Scale
VF000C7	VF	Dropped and moved to <i>Creating</i> Scale
VF000A8	VF	Non-cognitive item; not scaled
VF000B8	VF	Dropped and moved to <i>Creating</i> Scale

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

5.2.3 Estimation of Group Score Distributions

For the assessment sample, group score distributions were estimated for a single Visual Arts proficiency scale using the program BGROUP. As with scaling, student weights were used at this stage of the analysis. Instead of using the actual background variables for the estimates (as had been done prior to the 1990 assessment), principal components of the background variables were used. The principal components employed accounted for 90 percent of the variance of the original background variables (as coded in Appendix G) selected for estimation. Principal components were employed to remedy problems of extreme collinearity among some of the original variables.

Research based on data from the 1990 Trial State Assessment suggests that results obtained using the 90 percent subset of components will differ only slightly from those obtained using the full set (Mazzeo, Johnson, Bowker, & Fong, 1992). Table 5-4 lists the number of principal components, as well as the proportion of scale score variance accounted for by the estimation model for grade 8. A list of the variables defining group membership can be found in Appendix D.

Table 5-4
NAEP 1997 Arts Assessment in Visual Arts
Proportion of Scale Score Variance
Accounted for by the Estimation Model

Number of Contrasts ¹	Number of Principal Components ¹	Proportion of Scale Score Variance Accounted for by the Estimation Model
459	145	.61

¹ Excluding the constant term

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

The BGROUP program estimates distributions of scores by combining information from item responses (i.e., likelihood information) and background variables (i.e., prior information) of individuals.

5.2.4 Transformation of the Visual Arts Calibration Scale for Reporting

Since the 1997 Arts Assessment in Visual Arts was developed and scaled using within-grade procedures, and since there was no prior Arts Assessment in Visual Arts with a comparable framework to which it was being linked, a new reporting metric was adopted. The results are reported on 0–300 scale. The same convention was adopted for the Music and Theatre assessments. As is shown in Table 5-5, the mean of the Visual Arts scale was set at 150 for each grade, and the standard deviation at 35. If the achievement distribution were normal, we would expect this range (0–300) to cover about 99.998 percent of the distribution. Note that any transformed scale scores below 0 were censored to values of 0, and any transformed scale scores greater than 300 were censored to values of 300.

Table 5-5
NAEP 1997 Arts Assessment
Means and Standard Deviations
of All Five Plausible Values
for the Visual Arts Responding Scale

Mean	S. D.
150	35

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Scores on this transformed scale are used as a basis for making mean and percentile comparisons between subgroups for the NAEP Arts Assessment in Visual Arts. One should note, however, that comparisons from one assessment (e.g., Visual Arts) to another (e.g., Music) do not have meaning.

To put scores onto a scale with a mean of 150 and a standard deviation of 35, a linear transformation is applied to the scores. This means that each score was multiplied by one constant (A) and then added to another constant (B), as illustrated by the formula

$$\theta_{target} = A \times \theta_{calibrated} + B$$

where A and B are linear transformation constants. The values of A and B for the Visual Arts sample are given in Table 5-6.

Table 5-6
*NAEP 1997 Arts Assessment Transformation Constants
for the Visual Arts Responding Scale*

A	B
35.98	149.85

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

5.3 CALCULATING THE MEAN PERCENT-CORRECT SCALE FOR VISUAL ARTS CREATING ITEMS

Table 5-7 gives information similar to Table 5-1, but for the *Creating* items. These items require that students be involved in some sort of creative activity, such as modelling with clay. These items were distinguished from *Responding* items. *Responding* items were reported using an IRT scale. *Creating* items were reported using a percent of total possible score for each administered session type (see Section 5.3.1). Table 5-7 gives item statistics for the blocks made up exclusively of *Creating* items. This table shows that mean item scores were quite similar across blocks VG–VI, indicating that the blocks are of similar difficulty. However, alpha reliabilities and average r-Polyserials varied quite a bit.

Table 5-7
*NAEP 1997 Arts Assessment Descriptive Statistics
 for the Visual Arts Creating Items by Block*

Statistic	VG	VH	VI
Number of Items in Scale	7	8	9
Unweighted Sample Size	688	696	736
Weighted Mean Item Score Averaged Over Block	.41	.41	.43
Weighted Alpha Reliability	.79	.67	.81
Weighted Average r-Polyserial	.76	.60	.69

NOTE: Some of the Arts Assessment items were scored several times to evaluate different aspects of responses. The items specified in this table include these separate scorings of responses as separate items.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Tables D-5 and D-6 in Appendix D list the interrater reliability information for *Creating* items in blocks VG–VI and the eight *Creating* items in blocks VC–VF, as well as for the constructed-response *Responding* items. All but two of the *Creating* items were polytomously-scored constructed-response items. The reliability, indicated by proportion exact agreement, for these polytomously scored items was usually above .80. The proportions ranged from .77 to .94 with an average of .85. Intraclass correlations (appropriate for polytomous items) were of similar magnitude, ranging from .73 to .98 with an average of .86.

5.3.1 The Percent-Correct Scale for *Creating* Items

As mentioned above, the *Creating* items were not scaled using IRT methods. These items were used to form a descriptive scale comprised of percent-correct scores of *Creating* items for each session type. A session is a combination of Visual Arts blocks administered to a subsample of students. Sessions were spiraled so that equivalent populations responded to each item. Table 5-8 shows the session structure of the Arts Assessment in Visual Arts.

Table 5-8
NAEP 1997 Arts Assessment
Session/Block Structure for Visual Arts

Session	<i>Responding</i> Blocks		<i>Creating</i> Blocks	Number of <i>Creating</i> Items
1	VE	VC	†	3
2	VC	VD	†	2
3	VD	VF	†	5
4	VD	VE	†	1
5	VF	VC	†	7
6	VE	VF	†	6
7	VD	VC	†	2
8	VF	VE	†	6
9	VE	†	VG	8
10	VE	†	VH	9
11	VE	†	VI	10
12	VC	†	VG	9
13	VC	†	VH	10
14	VC	†	VI	11
15	VD	†	VG	7
16	VD	†	VH	8
17	VD	†	VI	9
18	VF	†	VG	12
19	VF	†	VH	13
20	VF	†	VI	14

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Students in each session received a book containing at least one *Responding* block. (*Responding* blocks are also called A/B blocks.) Students in some sessions also received a *Creating* block. Note that all of the *Responding* blocks, except VD, contain some items that are considered *Creating* items. These items are not included in the IRT scale but *are* included in the percent-correct *Creating* scale. The fourth column of Table 5-8 shows the number of *Creating* items in a session, which is the sum of the *Creating* items from the *Responding* blocks (VC, VD, VE, VF) and the *Creating* items from the *Creating* blocks (VG, VH, VI). The *Creating* percent-correct score for a student is the total score of the student (for the session) divided by the maximum possible score (for the session) multiplied by 100. For example, in Session 1 there are three *Creating* items, and the maximum possible score on these items is 11. If a student attained a total score of 5 for the session, his or her percent-correct score would be $(5 \div 11) \times 100 = 45$. It should be noted that these scales are descriptive, as different students received different items of varying difficulty. However, since the sessions were spiraled, all relevant subgroups were equally exposed to all session types. The percent-correct scale allows for making meaningful comparisons among subgroups of students in this assessment. One item in each of the *Creating* blocks

(VG, VH, VI) was not included in the Visual Arts *Creating* scale. Table 5-9 shows the items that went into the Visual Arts *Creating* scale.

Table 5-9
NAEP 1997 Arts Assessment
Items Contributing to the Visual Arts Creating Scale

Block	Item
VC	VC000A6
VC	VC000B6
VE	VE00009
VF	VF000B6
VF	VF000C6
VF	VF000B7
VF	VF000C7
VF	VF000B8
VG	VG000A1
VG	VG000B1
VG	VG000C1
VG	VG000D1
VG	VG000E1
VG	VG000F1
VG	VG000G1
VH	VH00001
VH	VH000A2
VH	VH000B2
VH	VH000C2
VH	VH000A3
VH	VH000B3
VH	VH000C3
VH	VH000D3
VI	VI00001
VI	VI00002
VI	VI000A3
VI	VI000B3
VI	VI000C3
VI	VI000F3
VI	VI000G3
VI	VI000H3
VI	VI000I3

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Appendix A

NAEP 1997 ARTS ASSESSMENT SAMPLE DESIGN AND WEIGHTING PROCEDURES¹⁰

*Lucy M. Gray and Nancy W. Caldwell
Westat*

A.1 INTRODUCTION

The sample for the NAEP 1997 Arts Assessment was selected using a complex multistage sample design. The statistical bases of sampling and weighting procedures used in NAEP and described in this appendix are given in the *NAEP 1996 Technical Report* (Allen, Carlson, & Zelenak, 1999). These analyses serve as a basis for the results presented in the Visual Arts sections of the *NAEP Arts Report Card* (Persky, Sandene, & Askew, 1998). The 1997 Arts Assessment sampled students from selected schools within 52 selected geographic areas, called primary sampling units (PSUs), across the United States. The measure of size for each PSU was the 1990 total population from the U.S. Census.

The sample design had four steps in the selection process:

1. Selection of geographic PSUs (counties or groups of counties),
2. Selection of schools within PSUs,
3. Assignment of session types to schools, and
4. Selection of students for session types within schools.

A.2 PRIMARY SAMPLING UNITS

The basic PSU sample design for the NAEP 1997 Arts Assessment is a stratified probability sample with one PSU selected per stratum with probability proportional to total population. Each PSU meets a minimum size requirement and consists of a Consolidated Metropolitan Statistical Area (CMSA), a Metropolitan Statistical Area (MSA), a New England County Metropolitan Area (NECMA), a single county, or group of contiguous counties in the U.S. (including Alaska, Hawaii, and the District of Columbia). A total of 52 PSUs were selected for the sample.

A.2.1 Selection of the Certainty and Noncertainty PSUs

Ten certainty PSUs were identified that had the largest 1990 population counts. Six additional PSUs were selected systematically from the 12 next largest PSUs with probability proportional to the

¹⁰ Lucy M. Gray assisted in survey operations and field activities for the NAEP 1997 Arts Assessment, under the direction of Nancy W. Caldwell.

1990 Census population. Before sample selection, these 12 PSUs were sorted by region and then serpentine (in an alternating ascending and descending pattern) by size within region. This particular design feature derives from the fact that this 52 PSU design was based on the larger 94 PSU design used for main NAEP assessments (see Allen, et al., 1999, Chapter 3).

Thirty-six noncertainty strata were formed, and one PSU per pair was selected for the sample, again with probability proportional to the 1990 total population. The major strata were defined by the four regions (Northeast, Southeast, Central, West) and metropolitan status (MSA, non-MSA). Further stratification was achieved by ordering the noncertainty PSUs by several socioeconomic characteristics. The characteristics used to define strata were the percent minority population, the percentage change in the total population since 1980, the per capita income, the percent of persons age 25 or over with college degrees, the percent of persons age 25 or over who completed high school, and the civilian unemployment rate. Up to four of these characteristics were used in any one major stratum.

A.3 SELECTION OF SCHOOLS

A.3.1 Frame Construction

The school sampling frame for the NAEP 1997 Arts Assessment was the same sampling frame of schools as was used for NAEP in 1996, but restricted to the 52 sample PSUs, and to schools that contain the target grade (Grade 8). The list of schools was originally obtained from Quality Education Data, Inc. (QED).

For each school in the frame, estimates were made of the number of students who were eligible by grade. The QED file gives total enrollment, enrollment by grade, and the grade range for each school, thus providing the average enrollment per grade.

Unlike many other NAEP samples, there was no oversampling of nonpublic schools or high minority schools for the Arts Assessment sample.

A.3.2 Sample of Schools

A measure of size was assigned to each school according to the following scheme. Let A_i denote the number of eligible students estimated to be enrolled in the school. The measure of size given to each school was:

$$\begin{aligned}
 0.25 &= && \text{If } A_i < 6; \\
 A_i/20 &= && \text{If } 6 \leq A_i \leq 19; \\
 1 &= && \text{If } 20 \leq A_i \leq 25, \text{ the maximum sample size in any one school; and} \\
 A_i/25 &= && \text{If } 25 < A_i.
 \end{aligned}$$

A systematic sample of schools was drawn. Schools were first sorted by certainty/noncertainty PSU status. Certainty PSU schools were sorted by school type (public/nonpublic). Within public school type the sort was by region, type of locale, stratum, measure of size, and within nonpublic school type¹¹ the sort was by region, school type, stratum, measure of size. The sorts were done in a serpentine fashion. For example, for public certainty schools, stratum would be sorted from ascending to descending within adjacent levels of type of locale, and the measure of size would be sorted from ascending to descending within adjacent levels of stratum. The systematic sample was then drawn from this list using the probabilities dictated by the size measure A_i . A total sample of 225 schools was selected. No substitutes were used for the NAEP 1997 Arts Assessment.

A.4 ASSIGNMENT OF SESSIONS TO SCHOOLS

Sessions were allocated to cooperating and pending schools only. All schools were assigned to either Visual Arts (“V”) or Music (“M”); 125 schools to Visual Arts and 100 schools to Music. All schools offering the appropriate courses in Theatre were assigned a Theatre session (“T”). Theatre is the only subject for which a school might be assigned multiple sessions. Each school was assigned one session of 25 students for either Visual Arts or Music. If there were ten or more students eligible for Theatre that were not selected for Visual Arts or Music, all of those students were assessed in Theatre. If there were any students “left over” on the list after selecting the Visual Arts or Music sample, all of those students who were eligible for Theatre were assigned to the Theatre session.

There were 20 sessions for Visual Arts. Sessions V1 to V8 each had a sample size of 100 students, and sessions V9 to V20 each had a sample of 150 students, for a total sample size of 2,600 students for Visual Arts. There were 10 sessions for Music, with a total sample size of 2,000 students. Sessions M1 to M6 each had a sample size of 250 students, sessions M7 and M10 each had a sample of 167 students, and sessions M8 and M9 each had a sample of 83 students. Theatre had 10 sessions, with a total sample size of 2,000 students. Sessions T1 to T6 each had a sample size of 250 students, sessions T7 and T10 each had a sample of 167 students, and sessions T8 and T9 each had a sample of 83 students.

A.5 SAMPLING STUDENTS

The sample of students within sampled schools was systematically drawn from school-prepared lists of eligible students. Student Listing Forms (SLF) were prepared for each participating school in a given age/grade class; all enrolled students of the specified age and all others in the corresponding modal grade were to be entered on the SLFs.

The Arts Assessment consisted of two major components: a paper and pencil (or booklet based) component and a performance activity. Generally the paper and pencil component includes two cognitive blocks, an “A” block and a “B” block. The performance activity is referred to as a “C” block. For students in Visual Arts, the performance activity was administered to all students in the session at the

¹¹ The 1997 Arts Assessment private school type variable had four possible values: Catholic, Other Religious, Unaffiliated, and Other/Nonreligious.

same time, while for Theatre students the performance activity was administered to small groups of two to four students. The performance activity for Music students was done on an individual basis.

A.5.1 Within-School Sampling Rates

To obtain the sampling rates within school we have the following:

Let

M_A = Maximum allowable sample size from an individual school (=25);

G_i = Revised estimate of age eligibles for school i .

Then the sampling rate applied to the list of eligible students was given by

$$R = M_A/G_i \text{ if } G_i > M_A.$$

Otherwise, the sampling rate was 1.0.

To control the student sampling operations as closely as possible, Westat generated a Session Assignment Form (SAF) for each school where sampling was to be carried out. This computer-generated form specified:

- the types of sessions that were to be administered at the school,
- the line numbers (from the SLF) specifying the students to be drawn into the sample,
- the minimum and maximum number of students listed on the SLF that could be accepted without requiring revision to the within-school sampling rates, and
- special instructions as appropriate for the Teacher Survey.

A.5.2 Sample Selection

Generally, the district supervisor carried out the sampling of students a week prior to the assessment. Student Listing Forms (SLF) were prepared for the applicable age class in each participating school. All enrolled students of the specified age and all others in the corresponding modal grade were to be entered on the SLF in any order convenient to the school, or the school could produce a computer-generated list. Before carrying out the sampling, the district supervisor reviewed the form and made comparisons with other information in an effort to make sure that the list included all eligible students.

The sampling was carried out according to very specific instructions described in the supervisor's manual. The sampling statisticians were available by telephone to assist in the resolution of sampling problems and to generate revised SAFs when necessary.

A.6 EXCLUDED STUDENTS

Specific groups eligible for exclusion were:

- limited-English proficient (LEP) students,
- students identified as having behavioral disorders, and
- students physically or mentally disabled, including Educable Mentally Retarded (EMR), in such a way that they could not respond to NAEP exercises as they were normally administered.

A.7 SCHOOL AND STUDENT PARTICIPATION RATES

Table A-1 summarizes the weighted participation rates of sampled schools, by public/nonpublic status and by assessment: Visual Arts, Theatre, and Music (General, Self-Selected). In parentheses next to the participation rate is the number of participating schools. Please note that the sum of the schools in the total column is larger than the total number of participating schools, due to multiple subject assessments within schools.

Table A-1
NAEP 1997 Arts Assessment
Weighted Participation Rates of Sampled Schools
by Public/Nonpublic Status

Session	Public	Nonpublic	Total
Music, General	79.46% (84)	83.26% (14)	80.01% (98)
Music, Self-Selected	77.60% (80)	77.68% (13)	77.61% (93)
Theatre	69.47% (40)	40.00% (2)	67.16% (42)
Visual Arts	83.63% (116)	85.21% (12)	83.77% (128)

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table A-2 summarizes the weighted participation rates of sampled students, after removing excluded students and students that are not age-eligible. Once again, participation rates are shown for public schools and nonpublic schools for all of the components: Visual Arts (A/B blocks, C block), Theatre (A/B blocks, C block), and Music (A/B blocks, C block, Self-Selected).

Table A-2
NAEP 1997 Arts Assessment
Weighted Participation Rates of Sampled Students
After Removing Excluded Students

Session	Public	Nonpubl ic	Total
Music, A/B Blocks ¹	91.05%	93.53%	91.33%
Music, C Blocks ²	81.83%	85.59%	82.23%
Music, Self-Selected	100.00%	100.00%	100.00%
Theatre, A/B Blocks ¹	78.88%	93.01%	81.68%
Theatre, C Blocks ²	72.08%	90.91%	73.37%
Visual Arts, A/B Blocks ¹	89.93%	94.99%	90.39%
Visual Arts, C Block ²	88.52%	94.16%	89.04%

¹ A/B blocks are also called *Responding* blocks.

² C blocks are also called *Creating* and/or *Performing* blocks.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

A.8 OVERALL STUDENT PARTICIPATION RATES

The overall student participation rates are calculated by multiplying the school participation rates and the student participation rates listed in Tables A-1 and A-2. They are as follows:

Table A-3
NAEP 1997 Arts Assessment
Overall Student Participation Rates

Session	Public	Nonpubl ic	Total
Music, A/B Blocks ¹	72.35%	77.87%	73.07%
Music, C Blocks ²	65.02%	71.26%	65.79%
Music, Self-Selected	77.60%	77.68%	77.61%
Theatre, A/B Blocks ¹	54.80%	37.20%	54.86%
Theatre, C Blocks ²	50.07%	36.36%	49.28%
Visual Arts, A/B Blocks ¹	75.21%	80.94%	75.72%
Visual Arts, C Block ²	74.03%	80.23%	74.59%

¹ A/B blocks are also called *Responding* blocks.

² C blocks are also called *Creating* and/or *Performing* blocks.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

A.9 TEACHER QUESTIONNAIRES

The teachers of sampled students that were assessed in Theatre were asked to fill out a teacher questionnaire. Teachers were asked to complete the questionnaire so that information about teaching methods may be linked to student performance.

A.9.1 Weighting Procedures

A.9.1.1 Base Weights

The base weight assigned to a selected student is the reciprocal of the probability that the student was sampled for the session type to which he/she was assigned. That probability is the product of the following eight factors for Self-Selected Music students and students in any of the C blocks. The probability of selection for students in the A/B block excludes the C-block factor and Self Select Factor:

1. The probability that the PSU was selected (*PSUWT*);
2. The conditional probability, given the PSU, that the school was selected (*SCHWT*);
3. The conditional probability, given the school, that the particular session type (Visual Arts, Music, Theatre) was assigned to the school (*SESSWT*);
4. The conditional probability, given the particular session type was assigned to the school, that the student was selected for that session type (*STUSCHWT*);
5. The weight reflecting how the sampled students were assigned to subjects in schools in which Theatre sessions were conducted (*SPLWT*);
6. The year-round adjustment factor (*YRNDADJ*);
7. The C block Factor, which is used for analysis of data collected in C blocks (*CBLOCK*); and
8. The Self-Selected probability factor, applied to those Self-Selected Music students that had to be sampled due to the fact that the school had a large number of eligible Self-Selected Music students (*SSFACT*).

The base weight for a Self-Selected Music student or a student in one of the C blocks may be expressed as the following product.

$$BASEWGT = PSUWT \times SCHWT \times SESSWT \times STUSCHW \times SPLWT \\ \times YRNDADJ \times CBLOCK \times SSFACT$$

where the variables on the right side of the equation are defined above.

The base weight for the students in the A/B blocks may be expressed as the following product.

$$BASEWGT = PSUWT \times SCHWT \times SESSWT \times STUSCHW \times SPLWT \times YRNDADJ$$

The PSU weight, *PSUWT*, is the reciprocal of the probability of selection for the PSU. Of the 52 PSUs selected, 10 were certainty PSUs and have a PSU weight of one. For the remaining 42 PSUs, the probability of selection accounts for the sampling procedure used.

The school weight, *SCHWT*, is the reciprocal of the probability of selection of the school conditional on the PSU.

The session allocation weight, *SESSWT*, is the inverse of the probability that the particular session was allocated to the school. This is a function of the session type and the number of sessions allocated to the school. Session allocation weights were calculated separately for each session type. Within each subject, there is one weight for the A/B block and three weights for each of the different performance C blocks.

For assessed students, the student weight, *STUSCHW*, gives the inverse of student's probability of selection for the particular session to which he/she was assigned. This probability is the product of the within-school sampling rate and the proportion of the relevant eligible students assigned to the particular session type within the schools, as prescribed by the SAF. For the Visual Arts C blocks, the student base weight was multiplied by 52/36 to account for the fact that in certain sessions students were not assigned a C block.

The weight associated with the skip interval for assigning the sampled students to subjects is *SPLWT*. For Visual Arts and Music the appropriate value is one; for Theatre the appropriate value is $(1/[STUSCHWT-1])$.

The year round adjustment factor, *YRNDADJ*, adjusts for the fact that some schools are in session all year long. For schools that are not in session year round the factor is one; for schools that are in session year round the factor is $(1/[1-PCTOFF/100])$, where *PCTOFF* is the percentage of the year the school is not in session.

The C-block factor, *CBLOCK*, is the inverse of the probability that the student was in one of the C blocks.

The Self-Selected factor, *SSFACT*, is the inverse of the probability that the student in the Music session is in one of the Music Self-Selected blocks to be sampled.

A.9.1.2 Adjustment of Base Weights for Nonresponse

The base weight for a student was adjusted by two nonresponse factors: *SESNRF*, to adjust for school nonresponse and schools that did not conduct all of their assigned sessions (i.e., a session nonresponse); and *STUNRF*, to adjust for students who were invited to the assessment but did not appear either in the scheduled or a makeup session. Thus, the nonresponse adjusted weight for a student was of the form:

$$STUAWGT = BASEWGT \times SESNRF \times STUNRF$$

where

$BASEWGT$ = student base weight described in the previous section.

The nonresponse adjustment factors were computed as described below.

Session Nonresponse Adjustment ($SESNRF$). Sessions were assigned to schools before cooperation status was final. The session nonresponse adjustment was intended to compensate for session nonresponse due to refusing schools or individual sessions not conducted. These factors were computed separately by session and across all sessions within classes formed by the first three digits of PSU stratum, called subuniverse (formed by crossing the PSUs major stratum and the first socioeconomic characteristic used to define the final stratum). Occasionally additional collapsing of classes was necessary to improve the stability of the adjustment factors, especially for the smaller assessment components. Most classes needing collapsing contained small numbers of cooperating schools. Occasionally, classes with low response rates were collapsed.

The nonresponse adjustment factors were computed separately for each subject (Music, Visual Arts, Theatre) in the A/B block, and for each subject for the three separate C blocks. There was also a separate nonresponse adjustment for the two Music Self-Selected blocks. The nonresponse adjustment took place in two steps since specific subject sessions were not assigned to all schools. The first nonresponse adjustment was for each subject in the A/B block, then these nonresponse adjusted weights were used as part of the base weight calculation for the C blocks and the Self-Selected blocks.

In subuniverse s , the session nonresponse adjustment factor $SESNRF_{hs}$ for session h was given by

$$SESNRF_{hs} = \frac{i \in B_{hs} \cdot PSUWT_i \times SCHWT_i \times SESSWT_{hi} \times G_i}{i \in C_{hs} \cdot PSUWT_i \times SCHWT_i \times SESSWT_{hi} \times G_i}$$

where

$PSUWT_i$ = The PSU weight for the PSU containing school i ,

$SCHWT_i$ = The school weight for school i ,

$SESSWT_{hi}$ = The session allocation weight in school i for session h ,

G_i = The estimated number of age-plus grade-eligible students in school i (based on QED data),

Set B_{hs} = Consists of all in-scope originally sampled schools allocated to session h in subuniverse s , and

Set C_{hs} = Consists of all schools allocated to session h in subuniverse s that ultimately participated.

Student Nonresponse Adjustment (STUNRF). Student nonresponse adjustment factors were completed separately for each assessment for assessed students, and for excluded students.

For excluded students, the student nonresponse adjustment was made separately for classes of students based on subuniverse. For assessed students, the adjustment classes were based on subuniverse, achievement level (student born before October 1982, student born October 1982 or later), and race/ethnicity class (White, Asian/Pacific Islander, other). In some cases, nonresponse classes were collapsed into one to improve the stability of the adjustment factors, although sometimes classes with low response rates were collapsed. For each class c in assessment g in session h , the student nonresponse adjustment factor $STUNRF_{ghc}$ is computed by

$$STUNRF_{ghc} = \frac{\sum_{j \in A_{ghc}^*} BASEWGT_{ghj} \times SESNRF_{hj}}{\sum_{j \in B_{ghc}^*} BASEWGT_{ghj} \times SESNRF_{hj}}$$

where

$BASEWGT_{ghj}$ = The within-school student weight for student j in session h in assessment g ;

$SESNRF_{hj}$ = The session nonresponse adjustment factor for the school containing student j in session h ;

Set A_{ghc}^* = The students in class c who were invited (i.e., selected and not excluded) to assessment g in session h , or the students in class c who were excluded from any session in the school; and

Set B_{ghc}^* = The students in class c who were assessed in assessment g in session h , or the students in class c who were excluded from any session, and for whom an SD/LEP questionnaire was completed.

A.9.1.3 Variation in the Weights and Trimming for Outliers

The trimming process trims the weight of students from any school that contributes more than a specified proportion to the estimated variance of the estimated number of students eligible for assessment. The same technical procedure was used as in recent previous national NAEP assessments (see Allen, Carlson, & Zelenak, 1999).

The students in some schools were assigned extremely large weights for one of the following reasons:

- The school was predicted to have a small number of eligible students, but in fact had a large number,
- The presence of large schools in PSUs with small selection probabilities,
- Small schools that were sampled with low probabilities, or
- High levels of nonresponse coupled with low to moderate probabilities of selection.

To reduce the effect of large contributions to variance from a small set of sample schools, the weights of such schools were reduced (i.e., trimmed). The trimming procedure introduces a bias, but is expected to reduce the mean square error of sample estimates.

A.9.1.4 Poststratification

The final student weights for Music and Visual Arts were poststratified using control totals for all grade 8 students. Theatre student weights were not poststratified because we were unable to obtain appropriate control totals for this subject, since the population of inference is not the total grade 8 population.

The weights determined in the manner described in the preceding sections were adjusted by poststratification in order to reduce the sampling error of estimates relating to student populations that span several subgroups of the total population.

Fourteen poststratification cells were formed using race/ethnicity, region, and age as shown in Table A-4. For each cell, the poststratification factor is a ratio whose denominator is the sum of the weights (after adjustments for nonresponse¹² and trimming¹³) of assessed and excluded students, and whose numerator is an adjusted estimate, based on more reliable data, of the total number of students in the cell (given above).

Table A-4
NAEP 1997 Arts Assessment
Poststratification Totals

Poststratification Cell	Race	Ethnicity	Census Region	Age	Cell Total
1	White	Non-Hispanic	1	13	336,158
2	White	Non-Hispanic	2	13	468,558
3	White	Non-Hispanic	3	13	530,057

(continued)

¹² The adjustment of base weights for nonresponse is described in Section A.9.1.2.

¹³ The adjustment to weights due to the trimming process is described in Section A.9.1.3.

Table A-4 (continued)
NAEP 1997 Arts Assessment
Poststratification Totals

Poststratification Cell	Race	Ethnicity	Census Region	Age	Cell Total
4	White	Non-Hispanic	4	13	348,590
5	All race/ethnicities	Hispanic	All	13	282,896
6	Black	Non-Hispanic	All	13	290,639
7	Other	Non-Hispanic	All	13	102,937
8	White	Non-Hispanic	1	Not 13	162,576
9	White	Non-Hispanic	2	Not 13	252,397
10	White	Non-Hispanic	3	Not 13	293,923
11	White	Non-Hispanic	4	Not 13	149,341
12	All race/ethnicities	Hispanic	All	Not 13	183,440
13	Black	Non-Hispanic	All	Not 13	240,518
14	Other	Non-Hispanic	All	Not 13	68,690

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

A.9.1.5 Final Student Weights

NAEP estimates of student characteristics are based on final student weights, that is, the weight resulting after adjusting the student base weight for nonresponse, trimming, and poststratification. The student final weight, *FINSTUWT*, is given by

$$FINSTUWT = STUAWGT \times TRIMFCTR \times PSFCTR$$

where

FINSTUWT = Nonresponse adjusted student base weight,

TRIMFCTR = Trimming factor, and

PSFCTR = Poststratification factor.

A.9.1.6 Replicating the Weights

All of the above weighting steps were replicated. Replicate weights were created for use with the jackknife method of variance estimation (see Allen, Carlson, & Zelenak, 1999). There were 34 replicates.

Appendix B

NAEP 1997 ARTS ASSESSMENT FIELD OPERATIONS AND DATA COLLECTION¹⁴

*Lucy M. Gray and Nancy W. Caldwell
Westat*

B.1 INTRODUCTION

This appendix describes the field operations and data collection activities for the national assessment component of the 1997 National Assessment of Educational Progress (NAEP) Arts Assessment in Music, Theatre, and Visual Arts, conducted only at the eighth-grade level. (A Dance assessment was also considered but was not implemented for the grade 8 assessment.) Consistent with other main NAEP assessments, the 1997 Arts Assessment was based on probability samples of schools and students that allow for regional and national reporting only. Because the 1997 data collection was an "operational probe," smaller school and student sample sizes were used than in a full-scale NAEP assessment.

Throughout the assessments, NAEP guarantees the anonymity of participants, and student or teacher names are never recorded on assessment booklets nor removed from the schools. NAEP results are reported at the national level and by region of the country, or sometimes by state, but never by school district, school, or individual student. Only group statistics are reported, broken down by gender, race/ethnicity, and a host of variables that illuminate teachers' instructional practices.

The 1997 Arts Assessment was conducted in a sample of approximately 225 public and nonpublic schools located in 62 geographic areas called primary sampling units (PSUs) throughout the states and the District of Columbia. The PSUs were selected by Westat to represent the nation as a whole. The assessments in the arts, at the eighth-grade level, were conducted from March 24–May 9, 1997.

The inclusion of the arts in the 1997 NAEP marked the third time the disciplines of Music and Visual Arts have been assessed nationally. Music was first assessed in 1972 and again in 1978; Visual Arts in 1975 and in 1978. The information gleaned from an assessment of the arts on a national scale broadens the scope of knowledge about what American students know and can do in the major arts disciplines.

B.2 ORGANIZATION OF THE ARTS ASSESSMENT

Each sampled school with grade 8 was scheduled for *one* assessment session in *either* Music or Visual Arts. If the school offered a NAEP-eligible course in theatre, up to three sessions of Theatre were conducted in the school, depending on the number of "Theatre-eligible" students. Students were eligible for Theatre only if they were enrolled in grade 8 *and* if they would have accumulated, by the end of the 1996–1997 school year, at least 30 in-school class hours in drama, acting, creative dramatics, play production, and/or technical theatre class.

¹⁴ Lucy M. Gray assisted in survey operations and field activities for the NAEP 1997 Arts Assessment, under the direction of Nancy W. Caldwell.

The Arts Assessment consisted of two major components: a paper-and-pencil (or booklet based) component and a *Performing* (or hands-on) activity. The paper-and-pencil component typically included two cognitive blocks, an "A" block and a "B" block. The *Performing* activity is often referred to as a "C" block. These blocks are described briefly below:

- **Visual Arts/Music.** The Visual Arts paper-and-pencil assessment included various prompts such as postcards depicting works of art; in Music, the paper-and-pencil component used excerpts of audiotaped musical selections.

Performance activities in Visual Arts included tasks such as designing a tool for a specific purpose and, in Music, performing music along with an accompaniment on audiotape. Performance tasks in Music were audiotaped and products created as a result of the Visual Arts activity were either collected or photographed.

- **Theatre.** In Theatre, the paper-and-pencil components focused on sequences from dramatic productions.

Students sampled for *Performing* tasks in Theatre were assigned roles of characters in a script. Because *Performing* tasks in Theatre were videotaped, parental consent was required for the *Performing* tasks.

The paper-and-pencil component for each session required approximately 1½–2 hours. Performance activities varied in length from 20–60 minutes and were administered either to all students in a session at the same time (Visual Arts), to small groups of 2–4 students (Theatre), or to individual students (Music). Since all students in a session were administered the Visual Arts tasks as one group, the Visual Arts *Performing* component usually was a continuation of the paper-and-pencil session. By contrast, students in Music and Theatre were scheduled to return in small groups or individually to take part in the *Performing* tasks.

The Music and Theatre *Performing* tasks were scheduled to begin the afternoon of the paper-and-pencil assessment and continue into the next day. Depending on the number of students to be assessed, it took two or three days to complete all assessment activities in a school.

Earlier, on the sampling day, the school questionnaire and the SD/LEP questionnaire(s) were distributed to appropriate school staff for completion. Teacher questionnaires were distributed only to theatre teachers of students selected for the Theatre assessment. Completed questionnaires were collected on the assessment day.

In order to reduce the burden on the participating schools, national assessment field staff performed virtually all of the work associated with the assessments. Introductory contacts and meetings were held in the fall (1996) to enlist cooperation and explain the assessment procedures to district and school representatives and to set a mutually agreed-upon assessment date for each school. The assessment supervisor visited the school to select the sample of students a week or two before the assessment. The assessment sessions were conducted by national assessment field staff, called exercise administrators, under the direction of the assessment supervisor. At the conclusion of the assessment in a school, field staff coded demographic information on the booklet covers and shipped the completed materials to Pearson Educational Measurement, the processing contractor for NAEP.

Historically, a small proportion (less than 10%) of the sampled students have been "excluded" from NAEP assessment sessions because, according to school records, they are either students with disabilities or limited-English proficient students who the schools determined should not participate in

the assessment. More recently, especially with the passage of the Individuals with Disabilities Education Act¹⁵, increased attention has been given to including as many of these students as possible in NAEP sessions. NAEP has addressed these concerns and has used new "inclusion" criteria (updated after 1996) and offered accommodations for testing students with disabilities and/or limited English proficiency (SD/LEP). For the Arts Assessment in 1997, grade 8 students were offered the following accommodations, if needed: extended time, read aloud, small group or one-on-one administration, large-print, and/or use of a bilingual dictionary.

B.3 PREPARING FOR THE ASSESSMENTS

B.3.1 Gaining the Cooperation of Sampled Schools

The process of gaining cooperation of the schools selected for the NAEP assessments began in August 1996 with a series of letters and contacts with state and district-level officials. The National Center for Education Statistics (NCES) first sent each jurisdiction a letter announcing NAEP plans for the 1997 data collection. Westat then contacted the State Test Directors or NAEP State Coordinators in each sampled state to notify them of the districts and schools selected in their states.

School cooperation and scheduling information was transmitted to Westat using a computerized data management system developed for this purpose. In early February, 1997, schools were sent a letter confirming the assessment schedule along with parental information letters (if requested by the school) and other informational materials.

Table B-1 provides the schedule of field activities for the Arts Assessment.

Table B-1
NAEP 1997 Assessment Schedule of Field Activities

Year	Month(s)	Field Activities
1996	Mid-September	Letter from Pascal Forgione, Commissioner, National Center for Education Statistics, U.S. Office of Education to Chief State School Officers announcing the 1997 NAEP program. NAEP sent the following materials to State Test Directors: <ul style="list-style-type: none"> ▪ cover letter, ▪ listings of sampled schools, ▪ summary of school tasks, ▪ sample letter to districts, and ▪ mailing labels for districts (if requested).
	September 25–September 28	Scheduling Supervisors' Training Session

(continued)

¹⁵ The IDEA Amendments of 1997 (Public Law 105-17) focus on improving teaching, learning, and educational results for students with disabilities. IDEA '97 makes clear that students with disabilities must be included in general state and districtwide assessment programs, performance goals and indicators are to be developed for these students, and the performance of students with disabilities is to be included in reports to the public.

Table B-1 (continued)
NAEP 1997 Assessment Schedule of Field Activities

Year	Month(s)	Field Activities
1996	September 30–December 6	<p>NAEP Field Managers and Scheduling Supervisors performed the following tasks:</p> <ul style="list-style-type: none"> ▪ contacted the State Test Directors to discuss the NAEP assessment and the appropriate means to contact sampled districts and schools in the state, ▪ developed a preliminary schedule of assessments for schools in each region, ▪ contacted districts and schools to secure cooperation and set the schedule of sampling visits and assessment sessions, and ▪ transmitted school cooperation and schedule information to Westat.
1997	February	<ul style="list-style-type: none"> ▪ Westat sent the following materials to schools sampled for the arts: informational materials; a letter confirming the assessment date; parental information letters, if requested; and grade 8 schools were sent copies of the Student Listing Form (SLF).
	March 3–March 8	<ul style="list-style-type: none"> ▪ Arts Assessment Supervisors’ Training.
	March 10–March 21	<ul style="list-style-type: none"> ▪ Arts supervisors called/visited schools to sample students, prepare assessment forms and questionnaires, and confirm final arrangements for the assessment.
	March 24–May 2	<ul style="list-style-type: none"> ▪ Arts Assessments conducted.
	May 5–May 9	<ul style="list-style-type: none"> ▪ Makeup sessions conducted as necessary.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

B.3.2 Supervisor Training

Training for assessment supervisors was multiphased and involved separate sessions that focused on gaining cooperation or sampling and data collection activities. All training was conducted by the Westat project director, field director, and home office staff. Also in attendance were representatives from Educational Testing Service (ETS), Pearson Educational Measurement, and NCES.

The first training session was held September 25–28, 1996 for field staff assigned to the gaining cooperation phase of the project. After an introduction to the study, which included the background and history of NAEP, an overview of the Arts Assessment, and the 1997 assessment schedule, the training continued with a thorough presentation of NAEP contact/gaining cooperation activities. This is a lengthy process of contacting states, districts, and schools regarding their participation in and scheduling for NAEP. Several demonstration phone calls, role-playing, and exercises were used to provide some practical experience during this part of the training. Other training topics included: supervisory responsibilities, setting the assessment schedule, recruiting and training exercise administrators, and

administrative forms and procedures. The scheduling supervisors also received a full day of training on using the reporting system installed on the laptop computers assigned to each of them for the gaining cooperation/scheduling phase.

The NAEP supervisors who were responsible for data collection activities were trained again, in a second session, held March 3–8, 1997. The training began with a review of the preliminary activities during the fall, including results of initial contacts with districts and schools, scheduling of assessments, the status of exercise administrators' recruitment. The main focus of the training was a thorough discussion of assessment activities, sampling procedures, inclusion of SD/LEP students, teacher surveys, providing testing accommodations, conducting the sessions, and administrative forms and procedures. Westat's classroom management videotape was also shown at this training session. Major portions of the March training were devoted to careful presentation of procedures involved in conducting the *Performing* activities, or "C" blocks, for the Arts Assessment. This entailed training in the use of numerous individual scripts associated with the *Performing* tasks and the operation of video and electronic equipment needed for the C blocks.

The assessment field managers were present at the various training sessions to support training activities and answer questions concerning participation in specific districts and schools. Each supervisor also met with their field manager and the person who completed the scheduling in their area, as a first step in preparing for the new supervisors' contacts with each school (and district, if needed).

B.3.3 Contacting Districts and Schools

Once the supervisors were trained in September 1997, they began working on obtaining cooperation. Westat supervisors gained cooperation from all sampled schools, including nonpublic schools in NAEP.

In previous NAEP assessments, the supervisors offered and usually held "introductory meetings" with representatives from the superintendents' offices and the selected schools, typically the superintendent and the principals. These served as both an introduction to NAEP and a presentation on what would be asked of the school. The meetings were also used to establish a schedule for the sampling visits and the assessments in the schools.

However, over the years, these meetings have become somewhat redundant since many districts have fallen into the NAEP sample more than one time. It has also become more and more difficult to schedule these meetings, as district and school officials find it harder to allot time away from their offices. Thus, during the preparations for 1997 NAEP, the material was almost always presented to the superintendents and principals during telephone calls rather than in formal meetings. Generally, only if an in-person meeting was specifically requested by the district or school officials, or if the supervisor felt that there was a better chance of convincing a district to participate in person, was such a meeting held.

As the supervisors contacted superintendents, principals, and nonpublic school officials to introduce NAEP and determine the schools' cooperation status, they completed two forms and entered the school status in the receipt control system installed on their laptop computers. The Results of Contact Form was completed to document the discussion the supervisor had with each administrator concerning the district's willingness to participate and any special circumstances regarding the schools' cooperation or assessments.

The supervisor also completed portions of a School Control Form. This form was preprinted with the number and types of assessment sessions assigned to the school, so that this information could then be shared with the district/school official. Information gathered during the phone call, including the name

of the person designated to be the school coordinator, the number of students in the designated grade, tentative dates for the sampling visit and assessment, and other information that could have some bearing on the assessment, was recorded on the form. This information was used to update records in the home office. In December, the forms were provided to the supervisors who would be conducting the assessments.

During the gaining cooperation phase, the supervisor discussed arrangements for the assessments with representatives from each school. Within the weeks scheduled for each PSU, the supervisor had the flexibility to set each school's assessment date in coordination with school staff. The staff sometimes expressed preferences for a particular day or dates or had particular times when the assessment could not be scheduled. Their preferences or restrictions depended on the events that had already been scheduled on their school calendar. Using this information from the schools, the supervisors set up the assessment schedule for each PSU. Careful planning and discussions were important in executing this task due to the complex nature of the Arts Assessment and the fact that, in many schools, 2–3 days were required to complete the *Performing* tasks.

The supervisor usually learned during the introductory contact whether a school required some form of parental notification or permission. Three versions of standard NAEP letters were offered for the school's use, and each letter could be produced for selected students only or for all eligible students. The first version informs parents about the assessment. The second assumes parental consent unless parents send the form back stating that they do not want their child to participate in the assessment. The third version requires that parents sign and return the form before students can be assessed. All versions of the letter were available to the schools, although when the issue of parental permission came up in a discussion, supervisors suggested using the least restrictive version that met the requirements of the school or district. Schools could also send out their own letters and notices if they preferred not to use those offered through NAEP. Information on whether the school required parent letters and the type of letter used was recorded on the School Control Form.

B.3.4 Recruiting, Hiring, and Training Exercise Administrators

During the fall, while the supervisors were contacting schools and scheduling assessments, another major task was to recruit and hire exercise administrators who would administer the assessment sessions and *Performing* tasks. Exercise administrators (EAs) were recruited from many sources. People who had served as exercise administrators before, with good evaluations from their previous supervisors, were usually the first considered for hiring, particularly if their backgrounds included skills and training in the arts. In many cases, specialized EAs, who were actually arts instructors or performers, were hired because of the special demands associated with administration of the arts sessions and *Performing* activities. Most of these EAs were identified through local newspaper ads and extensive telephone and personal interviews.

The assessment supervisors, all of whom were experienced NAEP supervisors, had complete responsibility for hiring and training the exercise administrators who would report directly to them. The training was standardized so that all supervisors used a prepared script and exercises to train the exercise administrators. These trainings required longer than the half-day sessions typical in NAEP due to the complexity of the Arts Assessment and associated materials.

Each exercise administrator received an exercise administrator manual, which covered the full range of their job responsibilities. After studying the manual, they attended the training session. During the training, the supervisor reviewed all aspects of the exercise administrators' job including preparing materials, booklets, and Administrator Schedules for assessments; the actual conduct of the session; post-assessment collection of materials; coding booklet covers; recordkeeping; and administrative matters.

Each exercise administrator was responsible for assisting the field supervisor in administering the individual *Performing* tasks (in addition to the pencil-and-paper sessions), so that a key (and time-consuming) part of their training involved learning the administration of numerous individual *Performing* tasks and the operation of multiple video and electronic equipment items used to conduct or record the C block tasks.

B.4 SELECTING THE STUDENT SAMPLES

After securing cooperation from the school, the first scheduled visit to each school was made to select the sample of students to take part in the assessments, and to conclude the arrangements for the actual testing. This visit was made in March 1997 by the supervisor responsible for the assessments in the school. Upon arriving at the school, the supervisor first reviewed the list of grade-eligible students and confirmed with the school coordinator that all eligible students were listed. If any eligible students were omitted, sampling could not proceed until the list was completed.

Using the computer-generated Session Assignment Form (SAF) for the main assessment, which was specific to the school, the supervisor selected the sample of students to be assessed in either Music or Visual Arts. The SAF documented the types of sessions to be administered, the anticipated number of students to be assessed, the expected number of students eligible for the assessment, and a series of line numbers designating the students to be sampled. After making sure that all eligible students had been listed, the supervisor numbered the students on the master list. If the total number of eligible students was within the minimum and maximum limits indicated on the SAF, the supervisor could proceed to select the sample. If the number was outside the limits, the supervisor called Westat for additional sampling instructions. With either the original instructions or revised line numbers, the supervisor proceeded to select the sample of students. The SAFs provided step-by-step instructions for sampling, indicating not just the line number of each student to be selected, but the type of assessment session for which each student was selected. Those eligible students on the school's master list whose line numbers were specified on the SAF were selected for the assessment.

For Theatre assessments at grade 8, NAEP's primary objective in sampling students was to select and assess *all* theatre students if feasible and acceptable to the school. Thus some schools may have had three or more Theatre sessions. Only students "eligible" for Theatre were selected and assessed, and this was defined as students enrolled in grade 8 who would accumulate, by the end of the 1996–1997 school year, at least 30 in-school class hours in one or more of the following theatre courses: Drama, Acting, Creative Dramatics, Play Production, and/or any Technical Theatre class. To select these students, the number of students eligible and *not* already assigned to another session type were counted and one of two specific sampling rules was applied depending on whether this count was smaller than 10 students or 10 or more students.

The first of these rules specified that if 10 or more students were eligible for Theatre and were not selected for the other sessions, sample and assess *all* of these students in Theatre unless the total eligible was unusually large and the school refused to assess more than approximately 90 students. If the school refused to assess all students, it was necessary to call Westat to request line numbers for the Theatre sample. If more than one Theatre session occurred in a school, the eligible students were renumbered and approximately equal numbers of students were assigned to each session type.

Otherwise, if fewer than 10 students (or *no* students) were eligible for the Theatre session, but some students assigned to the other session were *also* eligible for Theatre, these students were assessed in *both* subjects if the school was agreeable. If the school did not agree, the students were kept in the session for which they were originally selected, and the Theatre session was conducted with as small a number as one student.

NAEP assigned students to more than one session if it was acceptable to the school as noted above. If some students were eligible for more than one type of arts session (e.g., Music and Theatre), these students were assigned to multiple sessions. For these students, a Special Situation Form was completed to document this "double sampling," and the booklet numbers used in *both* sessions were recorded on the form.

Once students were assigned to sessions, the supervisor and exercise administrators filled out an Administration Schedule for each session. The Administration Schedule is the primary control document for the assessment. It is used to list each sampled student and is the only link between booklets and students. The sample was designed so that about 30 students were assigned to each session. The location of the session was filled in on the Administration Schedules. Arrangements for the additional time to conduct the *Performing* block activities, for individual students or small student groups, were also discussed thoroughly with the school coordinator. Because student names were recorded on the Administration Schedules, those forms remained in the schools after the sample was drawn.

The supervisor then asked the school coordinator to identify any students in the sample with an Individualized Education Program (IEP) (for reasons other than being gifted and talented) and/or who were designated as LEP. Any student with either (or both) of these designations was to be indicated on the Administration Schedules. The school was asked to complete an SD/LEP questionnaire for each student with this designation. This was to be completed by a teacher, counselor, or other school official who knew the designated student well.

The school coordinator was also asked to determine whether any of these students should be excluded from NAEP based on the criteria for assessing SD/LEP students. If the school coordinator could not identify the excluded students while the supervisor was at the school, the instructions were left with the coordinator along with blank copies of the SD/LEP questionnaire. In those cases, the coordinator consulted with other school officials and informed the supervisor as to who was to be excluded when he/she returned for the assessment.

At the end of the sampling visit, if requested by the school, the supervisor and/or exercise administrators made lists of the sampled students for the teachers and/or completed appointment cards notifying students about their assessment schedule. Teacher notification letters were also prepared in some schools, which explained the assessment and listed the students who had been selected.

B.5 CONDUCTING THE ASSESSMENT SESSIONS

B.5.1 Paper-and-Pencil Assessments

The primary responsibility for conducting the A/B block, paper-and-pencil assessment sessions was given to the exercise administrators. Supervisors were required to observe the first session each exercise administrator conducted to ensure that they followed the procedures properly. Supervisors were also required to be present in all schools with more than one small session to be conducted. The supervisor plays an important role as the liaison between the national assessment and school staff ensuring that the assessments go smoothly.

To ensure that sessions were administered in a uniform way, the exercise administrator was provided with scripts for each session type. The scripts were read verbatim. The scripts began with a brief introduction to the study. The exercise administrator then distributed the booklets, being careful to match the student with the preassigned booklet.

After the booklets were distributed, some additional, scripted directions were read. Students were asked to write in the NAEP school ID, and given some general directions in completing the assessment. For eighth-grade students, the first item of the background questions was read aloud by the exercise administrator, and the students read the rest of the items to themselves. Other than during the background questions, the students were told that any questions they might have could not be answered by the exercise administrator.

During the sessions, the exercise administrators walked around the room monitoring the students to make sure they were working in the correct section of their booklet and to discourage them from looking at a neighbor's booklet.

At the end of each assessment session, booklets were collected and students dismissed according to the school's policy. The exercise administrator was then responsible for completing the information at the top of the Administration Schedule, totaling the number of participating students, and coding the covers of all booklets, including those booklets assigned to absent students.

Also during the paper-and-pencil session, the NAEP supervisor spent some time in each assessment room to observe the session and to schedule all eligible students for the C block, or *Performing* task. Only students who participated in the A/B block session and (for Theatre sessions) had parental consent were eligible for the C block. Using the Administration Schedule on which attendance was recorded, a Scheduling Roster and student appointment cards were prepared for the *Performing* tasks, and the cards were distributed at the end of the paper-and-pencil (A/B) session. Any scheduling problems were resolved with individual students as they left the assessment room.

Included in this process, for Music sessions only, were special steps to identify students eligible for an advanced Music C-block activity. Before starting the paper-and-pencil session, students who were in a band or orchestra or chorus were asked to identify themselves and their instrument or voice part. While the students worked in their booklets, the information provided by the students was used to assign the "advanced" students to two *Performing* tasks, that is, to both a "regular" and an "advanced" task. If 10 or fewer students were eligible for an advanced session, all of these students were selected, or if more than 10 students were eligible, a sample of 10 "advanced" students was selected using a random number table. These students were scheduled for two *Performing* blocks, a C block and an Advanced block, to be conducted, consecutively, on the day after the paper-and-pencil session to allow them time to bring their instruments to the *Performing* session.

B.5.2 Performance Tasks

Conducting and administering the *Performing* tasks was the responsibility of the NAEP supervisor who always took the lead role in conducting these sessions. Several EAs were also used in each of these sessions to handle and operate the electronic equipment (boom boxes, video cameras, VCRs, etc.) and to demonstrate activities or conduct warm-up activities with the students.

All *Performing* tasks were administered in a uniform manner by following a specific script for each activity. Careful arrangements were made for appropriate space to conduct the tasks, and extraneous objects (e.g., yardstick, chalk, eraser) that might be "misused" by students during a *Performing* were removed from the room. Students were advised prior to the sessions to wear "appropriate" attire and that items such as skirts, high heels, and very tight or very loose clothing should be considered unacceptable for *Performing* tasks, particularly in Theatre. The NAEP staff set up and operated all video equipment, recorders, tape players and/or cameras needed to conduct and record each session/task according to documented specifications and extensive training received prior to the assessment period. Specifications for the exact placement and operation of the cameras and other equipment were provided in NAEP

assessment manuals. As each student entered the room for the *Performing* task, attendance was recorded on a Scheduling Roster and later transferred to the NAEP Administration Schedule to show that the student's C block had been completed.

Visual arts sessions differed from that of the other disciplines in that *Performing* tasks were, generally speaking, indistinguishable from most "paper-and-pencil" tasks. Depending upon the session type assigned, students may have been asked to create art objects of some sort throughout the assessment in both the paper-and-pencil session as well and in the C block session. The session script indicated the materials needed and any special procedures required in the distribution, collection, and labeling of materials. Most Visual Arts materials were packed by session to facilitate distribution. All student materials shipped back to Pearson Educational Measurement for scoring were labeled with the student's booklet ID number. Caution was used so that Visual Arts materials were never stored overnight in a location where the temperature was likely to fall below 32°F. Some Visual Arts sessions required students to create a three-dimensional object using modeling clay, and each student's sculpture was photographed from several angles (as specified in the script) at the end of the session, and the appropriate minilabel with the preassigned booklet ID was affixed to each photograph as well as to each student return envelope.

All **Music** *Performing* tasks were conducted on an individual basis and were audiotaped. The Music *Performing* tasks were administered by the person designated as the "Music EA." This session administrator was assisted by an EA who helped prepare the room, prepare and label the audiocassettes for each student, assemble the required stimulus tapes and sheet music in preparation for each assessment, and ensure that the next student was waiting to be assessed. A "boom box" was used to play all audiotaped stimuli and a Marantz tape recorder was used to record the session. An external microphone was used in conjunction with the Marantz to enhance the quality of the recording. During a *Performing*, the tape was never stopped for any reason unless specified in the script. Each student's *Performing* cassette was labeled with the student's booklet ID number. A separate audiotape cassette labeled with the student's booklet ID number was used to record the *Performing* tasks for students who also performed the advanced music study. These students were assigned a second *Performing* task "related to their specialty" and were asked to either sing or play their own musical instrument (or a keyboard provided by NAEP) according to "advanced" task specifications provided by NAEP for each school.

Theatre *Performing* tasks involved at least two students at a time depending on the activity. Each group's *Performing* task was recorded on a single videotape. For example, if the task required the *Performing* of one pair of students, only the group *Performing* of that pair was recorded on the videotape. Likewise, if four students were assessed together for a task, only the group *Performing* of those four students was recorded on the videotape. The Theatre warm-up and *Performing* tasks were conducted by NAEP staff. Students were asked to wear colored, numbered vests to distinguish one student from another in the video. The student booklet ID's were recorded on a Video Log along with school information and session information. The completed video tape for each session was labeled with session and booklet ID information.

B.6 RESULTS OF THE NAEP ARTS ASSESSMENT

B.6.1 School and Student Participation

The unweighted school response rate for eighth-grade Arts Assessments in 1997 was 82 percent overall. The final sample of cooperating schools included 226 schools with grade 8 arts classes, and 220 of these were public schools and 26 were nonpublic. The overall response rate of 82 percent is similar to the 1996 NAEP response rate for eighth-grade schools.

The overall student response rate for the 1997 Arts Assessment was similar (roughly 90%) to the rate for eighth grade in other NAEP assessments. The rate was lowest for Theatre assessments (83%) possibly because substantially fewer schools have theatre programs.

Similar to 1996, the school response rates for 1997 were somewhat lower than NAEP rates in other years. The most frequently stated reason for school and district refusals, historically, has been the increase in testing throughout the jurisdictions and the resulting difficulty in finding time in the school schedule to conduct the NAEP assessments. With so many states now mandating their own testing, school schedules are becoming tighter, and administrators are finding it increasingly difficult to accommodate outside testing. Despite the increased visibility and publicity surrounding NAEP, schools are reluctantly finding it necessary to decline participation as a result of the increasing demands on their students' time. In addition, the arts assessments are time-consuming and place a greater burden on schools, which can also impact the refusal rate.

Of the approximately 7,700 students sampled for the 1997 Arts Assessment, roughly three percent overall, were excluded by schools and roughly 10 percent either refused or were absent. Thus, approximately 6,600 students were assessed in one of the three session types: 2,265 students were assessed in Music, 1,390 were assessed in Theatre, and 2,974 students were assessed in Visual Arts.

The response rate at which supervisors were required to conduct a makeup session was set at 90 percent which is the same standard that has been used in recent main NAEP assessments. Only about 235 students were assessed in makeup sessions, and this boosted the overall student response rate by about 3 percentage points.

B.6.2 Assessment Questionnaires

Westat provided each school with a school questionnaire a few weeks before the assessment was scheduled to be conducted (i.e., at the time of sampling). At the same time, supervisors prepared an SD/LEP questionnaire for each sampled student with either an IEP and/or LEP designation, with the request that it be completed by someone at the school knowledgeable about that student. teacher questionnaires, *only* for eighth-grade *theatre* teachers, were also prepared at the time of sampling.

The NAEP supervisor requested that the various questionnaires be distributed as quickly as possible after the sampling so that they could be collected on the day of the assessment. If the questionnaires were not complete at the time of the assessment, the supervisor left a postage-paid envelope addressed to Pearson Educational Measurement to be used to return the questionnaires. Table B-2 summarizes the number of questionnaires distributed and the number completed.

Table B-2
NAEP 1997 Assessment Questionnaires Distributed and Completed

Questionnaires for Grade 8 Arts Assessment	School Questionnaire	Teacher Questionnaire (Theatre only)	SD/LEP Questionnaire
Number Expected (Distributed)	222	49	916
Number Received (Completed)	209	47	863
Percent Received	94%	96%	94%

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

B.7 FIELD MANAGEMENT

Two field managers monitored the work of about 25 scheduling supervisors who worked during fall 1996 to gain cooperation of districts and schools for the main assessment. During the assessment period, these staff were expanded to about 80 supervisors and 5 field managers. All supervisors reported directly to their field managers who, in turn, reported to Westat's field director. All contacts were made at least weekly.

An automated management system was developed and maintained in Westat's home office. The scheduling supervisors working to contact schools during the fall used this system on their portable computers. The system contained a record for each sampled school. A disposition code structure was developed to indicate the status of each school's participation (e.g., school cooperating, decision pending, school refusal, district refusal, school closed, etc.). As a school's status was determined, the scheduling supervisors entered the status of the school onto their computers, and this information was downloaded onto the home office system on a weekly basis. Disposition reports were then generated from the receipt system once a week so that home office staff could review the progress of securing cooperation from the sampled schools.

These reports were an invaluable tool for the sampling statisticians as well as for the field director and field management staff. They provided the statisticians with the information needed to determine whether the sample of schools was adequate to produce representative results. Based on the information contained in these reports, the sampling statisticians selected substitute schools to replace some of the noncooperating schools.

After assessments were completed, the system was used to enter data from the School Worksheets (for main NAEP) on the number of students to be assessed, the number assessed, and the number absent for each school. Data on completed questionnaires received was provided by Pearson Educational Measurement. The system was also used to alter school assessment dates, particularly when weather conditions required a change in schedule, and to monitor plans for and progress in conducting makeup sessions. Reports were generated weekly during the assessment period that allowed the project staff to monitor the progress of the assessments both in terms of checking that the schools were assessed on schedule as well as assuring that a high response rate was achieved. The sampling statisticians used these reports to monitor the sample yield by school, PSU, and age/grade level.

Progress of the assessments was constantly monitored through telephone reports held between NAEP supervisors, field managers, and home office staff. During these phone conversations, the supervisors' schedules were reviewed and updated, and any problems that the supervisors were experiencing were discussed. Progress of the fieldwork was also monitored during quality control visits made to the field by Westat and ETS office staff.

The supervisors who traveled filled out a Work Schedule for a one- to two-week period, showing their whereabouts, so that they could be contacted if necessary. It also allowed field managers and project staff to review the supervisors' schedules and the distribution of work.

Appendix C

NAEP 1997 ARTS ASSESSMENT PROCESSING AND SCORING ASSESSMENT MATERIALS¹⁶

Connie Smith
Pearson Educational Measurement

C.1 INTRODUCTION

Pearson Educational Measurement, under subcontract to Educational Testing Service (ETS), was responsible for the following tasks for both the eighth grade Arts Assessment and the twelfth grade Arts Field Test:

- printing of test booklets and questionnaires,
- materials packaging and distribution,
- receipt control,
- data capture through image, optical mark recognition scanning, and key entry,
- data editing and validation,
- professional scoring of constructed-response items and student *Performings*,
- data file creation, and
- inventory control and materials storage.

Pearson Educational Measurement received and processed a total of approximately 21,948 assessed student booklets and 1,119 questionnaires for the grade eight Arts Assessment and grade twelve Arts Field Test. A total of 404,914 readings of student responses were conducted via image and paper-based scoring. This allowed for item-by-item based scoring and online, real-time monitoring of inter-rater reliabilities, and the *Performing* of each individual reader. Scoring of the Arts included rating both constructed-response items and student *Performing* work which included evaluating original artwork, videotapes of Theatre and Dance *Performings*, and audio tapes of Music *Performings*. The tracking, linking, and logistical issues associated with data capture and *Performing* scoring of the many different components of the grade eight Arts Assessment and grade twelve Arts Field Test provided unique challenges and opportunities for all involved. The collaboration of all organizations involved in these two assessment programs was a critical part of their successful completion.

C.2 OVERVIEW

For the NAEP 1997 Arts Assessment (grade 8 only) and the grade 12 field test, 76 unique documents were designed. Pearson Educational Measurement printed more than 149,000 booklets and forms. Printing preparations began with the design of the booklet covers in the Summer of 1996. This was a collaborative effort involving staff from ETS, Westat, and Pearson Educational Measurement. Since the goal was to design one format for use with all of the documents, necessary data elements to be

¹⁶ Connie Smith supervised the processing and scoring of materials for the Arts Assessment.

collected for the various booklets had to be determined and their placement on each cover type had to be agreed upon. After various iterations, the cover design was finalized in November of 1996. In a similar collaboration with ETS and Westat, Pearson Educational Measurement prepared administration schedules and control documents for the Arts Assessments.

ETS created camera-ready blocks using Design Expert™ software for the field test booklets and questionnaires. This data was sent to Iowa City on SyQuest disks along with a paper version of each block showing the page layout of text and artwork as it would appear in the printed document. Copies were made of the paper version of each block to be used for Pearson Educational Measurement review. NAEP staff in Iowa City coordinated this review procedure and referred any questions and suggestions to the designated staff member at EST for resolution. Suggestions for changes to data at this stage were analyzed. If minor, the changes were made by Pearson Educational Measurement staff in Iowa City. Extensive changes were made by ETS and a revised version of the entire block was sent on disk to Iowa City. After a block had passed the quality checking, the camera-ready data on disk was sent to the printer in Columbia, PA along with a guide indicating the number of times each block would be repeated in the assessment battery.

The actual “building” of booklets began after all blocks needed for a particular booklet were received and the Office of Management and Budget (OMB) had given its approval to print. ETS supplied booklet maps, which specified the order of blocks in each booklet. Using these booklet maps and actual mock-ups of booklets as guides, the Pearson Educational Measurement printer assembled electronic components into complete booklets. Generally, four weeks elapsed between receipt of final copy and delivery of printed booklets. The printer forwarded proofs to ETS and to Pearson Educational Measurement for review and approval to print. After ETS approved these proofs, Pearson Educational Measurement communicated ETS approval to the printer, along with any changes or corrections that needed to be made.

Performance blocks for the fine arts, called C blocks, were treated as separate documents. Covers were created on a Macintosh computer, and the scanned version of the artwork was added to each cover. Eight C blocks were printed as "cover only." These covers accompanied *Performings* recorded on audio or video tape that required no additional written response. The remaining C blocks varied from two to ten pages in length and were received from ETS as camera-ready scannable pages in standard block format. These were modified by Pearson Educational Measurement to serve as stand-alone key-entry booklets. Halftones of photographic illustrations were prepared for proper reproduction. All C-block booklets were reproduced by the Pearson Educational Measurement Copy Center in Iowa City.

As booklets and forms were printed, pallets of documents were received and entered into the Pearson Educational Measurement Inventory Control System. Sample booklets were selected and quality checked.

C.3 PACKAGING AND DISTRIBUTION

The distribution effort for the NAEP 1997 Arts Assessment (grade 8 only) and the grade 12 field test involved packaging and mailing documents, associated forms, and arts materials to the Westat supervisors. The NAEP Materials Distribution System (MDS) was utilized again in 1997. Files in the MDS system contained scheduled assessment dates, the names and addresses for shipment of materials, and a listing of all materials available for use by a participant in a particular subject area. Changes to any of this information were made via file updates provided by Westat.

Bar code technology continued to be utilized in document control. Pearson Educational Measurement identified each document with a unique ten-digit ID number. This number consisted of the

three-digit booklet number or form type, a six-digit sequential number, and a check digit. Each form was assigned a range of ID numbers. Bar codes reflecting this ID number were applied to the front covers of documents by Pearson Educational Measurement bar code processes and high-speed ink jet printers.

Booklets for Dance, Music, Theatre, and Visual Arts were not spiraled, therefore once all booklets from a subject area were bar coded, they were bundled into groups of eleven documents.

Each bundle of eleven documents had a bundle slip/header sheet to indicate the subject area, bundle type, bundle number, and a list of booklet types to be included in the bundle. To ensure the accuracy of each bundle and the security of the NAEP assessment, a quality control plan was used to verify the document order of each bundle and to account for all booklets. All bundles that contained a bundle slip were taken to a bar code reader/document transport machine where they were scanned to interpret each bundle's bar codes. The file of scanned bar codes was then transferred from the personal computer connected to the scanner to a mainframe data set.

The unique bundle number on the header sheet informed the system program as to what type of bundle should follow. A computer job was run to compare the bundle type expected to the sequence of booklets that were scanned after the header. This job also verified that the appropriate number of booklets was included in each bundle. Any discrepancies were printed on an error listing and forwarded to the Packaging Department. The error was corrected and the bundle was again read into the system. This process was repeated until no more discrepancies existed.

Once all bundles for a subject area passed the bundle QC process, each bundle was shrink-wrapped and flagged on the system as ready for distribution. Information from the bundle QC file was then uploaded to the mainframe computer system and used in the creation of administration schedules and Student ID Labels. Administration schedules for each scheduled session were pre-printed with the booklet IDs designated for that session. Three bundles of booklets were preassigned to each session except for grade 8 Theatre sessions where only two bundles were assigned. This number of booklets, 33 for Dance, Music, Visual Arts, and grade 12 Theatre sessions and 22 booklets for grade 8 Theatre sessions, most closely approximated the average projected session sizes and allowed supervisors an additional supply of booklets for extra students.

Using sampling files provided by Westat, Pearson Educational Measurement assigned bundles to schools and customized the packing lists. File data from Westat was coupled with the file of bundle numbers and the corresponding booklet numbers. This file was then used to preprint all booklet identification numbers, school name, school number, and session type directly onto the scannable administration schedule. As a result, every prescheduled session had specific bundles assigned to it in advance. This increased the quality of the booklet accountability system by enabling Pearson Educational Measurement to identify where any booklet should be at any time during the assessment. It also eliminated the possibility of transcription errors by field staff who would otherwise hand write booklet ID numbers on the administration schedule.

Students participating in the Arts assessments were, in most cases, assigned two booklets. The first, or A/B, booklet contained cognitive blocks and was a scannable document. The second booklet assigned to the student, known as the C block, contained *Performing* tasks and was a key-entry document. To create a link between the bar coded A/B booklets and the non-bar coded C-block booklets, the same sampling files used to create preprinted Administration Schedules were used to create Student ID Labels. These Student ID Labels matched the booklet IDs preprinted on the Administration Schedule and the bar codes on the A/B booklet covers assigned to that session. Westat supervisors affixed the appropriate Student ID Labels to the C-block booklet that corresponded to a student's A/B booklet and to the artwork produced by that same student. Having the preprinted Student ID Labels helped further reduce the possibility of error in supervisor transcription of booklet IDs and reduced the amount of time

supervisors spent filling out tracking documents and booklet covers. At the same time, Student ID labels improved the quality of the matching link between a student's A/B booklet, C-block booklet, and artwork.

Distribution of materials was accomplished in two phases. In the first phase, bulk supplies of materials, which consisted of reusable items such as audio and video tapes, were distributed to each supervisor. The second phase was the distribution of session specific materials by Supervisor Region and Primary Sampling Unit (PSU). Each session box of materials contained the assigned bundles of booklets and the appropriate ancillary items.

Initially, a total of 661 individual sessions were shipped for the Arts Assessment and the grade 12 field test. An additional 245 shipments of booklets and miscellaneous materials were sent throughout the assessment period. In addition to preparing student booklets for distribution, Pearson Educational Measurement also assembled numerous art portfolios, kits, and print sets.

All outbound shipments were recorded in the Pearson Educational Measurement Outbound Mail Management system. Each box shipped contained a bar coded label which was read by a scanner. This label contained the supervisor's region and shipping address. This information was transferred from a PC and uploaded to the mainframe at the end of each day. A computer program could then access information to produce reports on all shipments sent. These reports helped Pearson Educational Measurement phone staff trace shipments not received by Westat supervisors.

If the quantities of materials initially shipped were insufficient to conduct a particular assessment, supervisors or Field Managers could request additional materials via the NAEP toll free line. For the first time, Pearson Educational Measurement also implemented the use of e-mail to request additional materials. The e-mail address was typically used by Field Managers procuring materials for either one or multiple supervisors. The NAEP toll-free line and e-mail address could also be used for general inquiries about the assessment, shipment tracing requests, and questions about initial shipment delivery dates.

To process a short shipment request, Pearson Educational Measurement phone staff asked the caller for the following information:

- school ID,
- PSU (primary sampling unit),
- supervisor's name,
- phone number (where supervisor can be reached),
- address to be sent to,
- materials needed, and
- materials due date.

This information was then entered into the online short shipment system. After the requested items, due date and method of shipment were entered, the system produced a packing list and mailing labels for Pearson Educational Measurement packaging staff.

Approximately 245 calls and e-mail requests were received regarding the NAEP Arts field test and Arts Assessment.

As mentioned previously, short shipment requests were also made via e-mail. When requests for additional materials were received via e-mail, Pearson Educational Measurement staff verified that all information pertaining to shipping addresses and dates of delivery were present in the e-mail, as well as the lists of materials being requested. Pearson Educational Measurement staff would verify receipt of the e-mail request by sending a confirmation e-mail to the originator of the request. The e-mail requests were made only for materials that would be needed by a supervisor at an advanced date. Any requests for materials that needed to be shipped overnight or 2-day delivery were made by phone.

C.4 PROCESSING OF TEST MATERIALS

This appendix describes the various stages of work involved in receiving and processing the documents used in the Arts Assessment and grade 12 field test. Pearson Educational Measurement staff created a set of predetermined rules and specifications for the processing departments within Pearson Educational Measurement to follow. Project staff performed a variety of procedures on materials received from the assessment administrators before releasing these materials into the Pearson Educational Measurement NAEP processing system. Control systems were used to monitor all NAEP materials returned from the field. The NAEP Process Control System (PCS) contained the status of sampled schools for all sessions and their scheduled assessment dates. As materials were returned, the PCS was updated to indicate receipt dates, to record counts of materials returned, and to document any problems discovered in the shipments. As documents were processed, the system was updated to reflect processed counts. Pearson Educational Measurement report programs were utilized to allow ETS, Westat, and Pearson Educational Measurement staff to monitor the progress in the receipt control operations. Report programs were dramatically enhanced for the 1997 assessments through a joint ETS, Pearson Educational Measurement, and Westat effort. These enhancements are described in detail in a later section. The number of booklets processed are shown in Table C-1.

Table C-1
NAEP 1997 Arts Assessment Participation Counts

	Total Assessed in Original Sessions	Total Assessed in Makeup Sessions	Number Absent	Total Assessed with Accommodations	Total
Music					
<i>A/B Blocks</i>	2,199	66	0	26	2,291
<i>C Block</i>	1,981	1	6	5	1,993
<i>Self-Selected Block</i>	620	0	35	0	655
Theatre					
<i>A/B Blocks</i>	1,329	61	1	8	1,403
<i>C Block</i>	1,220	6	0	2	1,228
Visual Arts					
<i>A/B Blocks</i>	2,900	74	0	49	3,023
<i>C Block</i>	2,056	40	0	32	2,128

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

An alert process was used to record, monitor, and categorize all discrepant or problematic situations. Throughout the processing cycle, alert situations were either flagged by computer programs or identified during clerical check-in procedures.

Certain alerts, such as missing demographic information on the administration schedule, were resolved by opening staff retrieving the information from booklet covers. These alerts, known as “Information Alerts,” were recorded directly into the PCS system by opening personnel, eliminating the need for paper documentation. Since these problem situations were categorized and tallied as they were key entered in the PCS system, project staff were able to provide timely reporting on clerical-type errors made during test administration.

Alert situations that could not be resolved by opening personnel were described on alert forms which were forwarded to project personnel for resolution. Once resolved, the problems and resolutions were recorded online by project staff in the PCS system.

Pearson Educational Measurement’s Work Flow Management System (WFM) was used to track batches of student booklets through each processing step, allowing project staff to monitor the status of all work in progress. It was also used by Pearson Educational Measurement to analyze the current work load, by project, across all work stations. Through routine monitoring of this data, Pearson Educational Measurement’s management staff was able to assign priorities to various components of the work and to monitor all phases of the data receipt and processing.

C.4.1 Document Receipt/Opening

Shipments were to be returned to Pearson Educational Measurement packaged in their original boxes. The bar coded label applied during the distribution phase containing the NAEP school ID number was scanned into a personal computer (PC) file upon receipt. The PC file was then transferred to the mainframe and the shipment receipt date was applied to the appropriate school within the PCS system, providing the status of receipts regardless of any processing delays. Each receipt was reflected on the PCS status report provided to the Pearson Educational Measurement receiving department and supplied to Westat via electronic file transfer and in hard copy format. ETS also received a hard copy. The shipment was then forwarded to the opening area.

Opening personnel checked the shipment to verify that the contents of the box matched the school and session indicated on the label. Each shipment was checked for completeness and accuracy. Any shipment not received within three days of the scheduled assessment date was flagged in the PCS system and annotated on the PCS Exceptions Report. The administration status of these delayed shipments was checked and in some cases a trace was initiated on the shipment.

Pearson Educational Measurement was required to open all shipments within forty-eight hours of their receipt and to key enter preliminary processing information into the PCS system from the administration schedule. The preliminary information was written on the administration schedule by Westat assessment administrators and consisted of the following:

- school number,
- session number,
- original test date,
- total number to be assessed,
- total number assessed, and

- completeness flag (if appropriate.)

This preliminary information, used to provide Westat with timely student response rates, was updated with actual data when materials passed successfully through processing error free. A completeness flag was also applied to the PCS file by Pearson Educational Measurement opening staff if any part of the shipment was missing or if a problem alert was recognized during the opening process.

If multiple sessions were returned in one box, the contents of the package were separated by session. The shipment was checked to verify that all booklets preprinted or handwritten on the administration schedule were returned with the shipment and that all administration codes matched from booklet covers to the administration schedule. If discrepancies were discovered at any step in this process, the receiving staff issued an alert to facilitate tracking. If the administrator indicated that a makeup session was being held, the documents were placed on holding carts until the makeup session documents arrived. If no makeup session was indicated, Westat was contacted for the status of the missing materials. If the missing materials were to be returned, the documents already received were held until that time. If the materials were not being returned, processing continued and the appropriate administration code was applied to the administration schedule.

C.4.2 Batching of Booklets

Once all booklets listed on the administration schedule for a session were verified as being present, the entire session (both the administration schedule and booklets) was forwarded to the batching area. The booklets were batched by grade level, subject, and session type. Each batch was assigned a unique batch number. This number, created on the Image Capture Environment (ICE) System for all Image Scannable documents and on the Work Flow Management (WFM) System for all Key Entry and OMR Scannable documents, facilitated the internal tracking of the batches and allowed departmental resource planning. All other scannable documents (school questionnaires, teacher questionnaires, SD/LEP questionnaires, and the Roster) were batched by document type in the same manner.

When batching Arts documents, Pearson Educational Measurement needed to allow for having both Image Scannable and Key Entry documents present in the same session, or having booklets listed on the administration schedule that would not be present in processing. This was due to the testing accommodation of large-print booklets and the use of C-block booklets which were key entry documents.

When batching Arts documents, it was critical to maintain control of the placement of documents within batches to facilitate the flow of paper scoring materials and to be able to locate the artwork associated with a particular session, whether it belonged to the A/B booklet or the C-block booklet portion of a given student's responses. This was accomplished through parallel batching.

With parallel batching, the image scannable A/B booklets from a session or group of sessions were assigned a batch number on the ICE System and given an "E" batch number prefix, while the key entered C-block booklet batch from the same session or group of sessions was created on the WFM System and assigned the same batch number using a prefix of "T." The A/B booklet batch was then forwarded to Image Scanning with the administration schedule used as a scannable session header. The C-block booklet batch was assigned a Key Entry session header containing the school ID number and session code found on the administration schedule and forwarded to the Key Entry department for processing. Once processing was complete for both batches, session information was linked by a computer program.

C.4.3 Batching and Scanning of Questionnaires

The Arts Assessment and the grade 12 field test used one roster to account for all questionnaires. The Roster of Questionnaires recorded the distribution and return of Students with Disabilities/Limited-English Proficient Students (SD/LEP) questionnaires, teacher questionnaires, and school questionnaires. Some questionnaires may not have been available for return with the shipment. These were returned to Pearson Educational Measurement at a later date in an envelope provided for that purpose. The questionnaires were submitted for scanning as sufficient quantities became available for batching. Batches of school questionnaires and Rosters, which were image scannable documents, were created on the ICE system. Teacher and SD/LEP questionnaire batches were created on the WFM system since these documents were OMR scannable. Batches were then forwarded to scanning

C.4.4 Booklet Accountability

In 1997, Pearson Educational Measurement used a sophisticated booklet accountability system to track all distributed booklets. Prior to the distribution of NAEP materials, unique booklet numbers were read by bundle into a file. Specific bundles were then assigned to particular supervisors or schools. This assignment was recorded in the NAEP Materials Distribution System. When shipments arrived at Pearson Educational Measurement from the field, all used booklets were submitted for processing and a “processed documents” file was maintained. Unused booklets were submitted for security scanning where booklet ID bar codes were read and recorded into a separate file. This file and the “processed documents” file were later compared to the original bundle security file for individual booklet matching. If intact bundles of unused booklets were returned from the field only the bar coded bundle slip identifying the booklets contained in the bundle needed to be scanned, eliminating the need to scan each booklet in the bundle individually. Batches of bundle slips were created and the output files from these batches were compared to the “processed documents” file in the same manner as files created from batches of security scanned booklets. All unused materials received were then inventoried and sent to the Pearson Educational Measurement warehouse for storage while awaiting authorization from ETS to salvage them.

C.4.5 Data Transcription

The transcription of the student response data into machine-readable form was achieved through the use of the following three separate systems:

- data entry (which included OMR and image scanning, intelligent character recognition [ICR], and key entry),
- data validation (edit) , and
- data resolution.

Pearson Educational Measurement used the same format as in prior NAEP assessments and field tests to set up the document definition files for the large numbers of unique documents used in the Arts Assessment and the grade 12 field test. To do the proper edits, a detailed documents definition procedure was designed to allow Pearson Educational Measurement to define an item once and use it in many blocks and to define a block once and use it in many documents.

C.4.5.1 Data Entry

The data entry process was the first point at which booklet-level data were directly available to the computer system. Depending on the NAEP document, one of three methods was used to transcribe NAEP data to a computerized form. The data on scannable documents were collected using Pearson Educational Measurement optical-scanning equipment which also captured images of the constructed-response items and intelligent character recognition (ICR) fields. In all three cases, the data were edited and suspect cases were resolved before further processing.

C.4.5.2 OMR Scanning

The data values captured from booklets, questionnaire covers, and administration schedules were coded as numeric data. Unmarked fields were coded as blanks and the editing staff was alerted to missing or uncoded critical data. Fields that had multiple marks were coded as asterisks (*). The data values for the item responses and scores were returned as numeric codes. The multiple-choice single response format items were assigned codes depending on the position of the response alternative; that is, the first choice was assigned the code “1,” the second “2,” and so forth. The mark-all-that-apply items were given as many data fields as response alternatives; the marked choices were coded as “1” while the unmarked choices were recorded as blanks.

C.4.5.3 Image Scanning

The images of constructed-response items were saved as a digitized computer file. The area of the page that needed to be clipped was defined prior to scanning through the document definition process. The fields from unreadable pages were coded “X” as a flag for resolution staff to correct. Any image document or sheet unreadable by the image scanning system was taken to a flatbed scanner to be scanned into the system. In addition to capturing the student responses, the bar code identification numbers used to maintain process control were decoded and transcribed to the NAEP computerized data file.

C.4.5.4 Intelligent Character Recognition

The Intelligent Character Recognition (ICR) engine was again utilized to read various hand and machine printing on the front cover of the assessment and supervisor documents for the Arts Assessment and the grade 12 field test. Some information from student documents, administration schedules, rosters of questionnaires, and some questions in the school questionnaires were read by the ICR engine and verified by an online key entry operator. In all, the ICR engine read 1,994,416 characters for the NAEP 1997 assessments. Use of the ICR engine saved NAEP field staff a significant amount of time since they did not have to grid rows and columns of data.

The scanners and programs were also able to read imprinted codes, known as 2-out-of-5 codes, that were printed on the administration schedule. These 2-out-of-5 codes were imprinted at the same time the booklet ID numbers were printed on the administration schedule, and identified which booklet ID's were listed on the document. This eliminated a significant amount of online editing time needed to process the NAEP assessments. When the scanning programs were unable to translate the 2-out-of-5 codes, image clips of the booklet ID numbers were displayed to online editing staff for verification.

C.4.5.5 Key Entry

A process of key entry and verification was used to enter and make corrections to the nonscannable C-block documents. teacher questionnaire and SD/LEP questionnaire information was also corrected using key entry methods. Pearson Educational Measurement used the Falcon system to enter this data. The terminal screens were designed to enhance operator speed and convenience. The fields to be entered were titled to reflect the actual source document. Therefore, all key entry fields were specific to the NAEP student documents or questionnaire types being keyed.

C.4.5.6 Data Validation (Editing) and Resolution

Each data set produced by the scanning system contains data for a particular batch. These data had to be validated (or edited) for type and range of response. The data-entry and resolution system used was able to simultaneously process a variety of materials from all age groups, subject areas, control documents, and questionnaires as the materials were submitted to the system from scannable and nonscannable media.

The data records in the scan file were organized in the same order in which the paper materials were processed by the scanner. A record for each batch header preceded all data records for that batch. The document code field on each record distinguished the header record from the data records.

When a batch header record was read, a preedit data record and an edit log entry was generated. As the program processed each record within a batch from the scan file, it wrote the edited and reformatted data records to the preedit file and recorded all errors on the edit log. The data fields on an edit log record identified each data problem by the batch sequence number, booklet serial number, section or block code, field name or item number and data value. After each batch had been processed, the program generated a listing or online edit file of the data problems and resolution guidelines. An edit log listing was printed at the termination of the program for all nonimage documents. Image “clips” requiring editing were routed to online editing stations for those documents that were image scanned.

As the program processed each data record, it first read the booklet number and checked it against the session code for appropriate session type. Any mismatch was recorded on the error log and processing continued. The booklet number was then compared against the first three digits of the student identification number on the administration schedule. If they did not match, a message was written on the error log. The remaining booklet cover fields were read and validated for the correct range of values. The school codes had to be identical to those on the PCS record. All data values that were out of range were read “as is” but were flagged as suspect. All data fields that were read as asterisks (*) were recorded on the edit log or online edit file.

The blocks in a document were transcribed in the order that they appeared in the document. Each block’s fields were validated during this process. If a document contained suspect fields, the cover information was recorded on the edit log along with a description of the suspect data. The edited booklet cover was transferred to an output buffer area within the program. As the program processed each block of data from the data set record, it appended the edited data fields to the data already in this buffer.

The program then cycled through the data area corresponding to the item blocks. The task of translating, validating, and reporting errors for each data field in each block was performed by a routine that required only the block identification code and the string of input data. This routine had access to a block definition file that had, for each block, the number of fields to be processed, and, for each field, the field type (alphabetic or numeric), the field width in the data record, and the valid range of values. The

routine then processed each field in sequence order, performing the necessary translation, validation, and reporting tasks.

The first of these tasks checked for the presence of blanks or asterisks (*) in a critical field. These were recorded on the edit log or online edit file and processing continued with the next field. No action was taken on blank fields for multiple-choice items. Since the asterisk code indicated a double response, these items were written to the edit log for possible resolution by editing staff. Each field was then validated for range of response and any values outside of the specified range were recorded on the edit log or online edit file. The program used the item-type code to make further distinction among constructed-response item scores and other numeric data fields.

Moving the translated and edited data field into the output buffer was the last task performed in this phase of processing. When the entire document was processed, the completed string of data was written to the data file. When the program encountered the end of a file, it closed the dataset and generated an edit listing for nonimage and key entered documents. Image scanned items which required correction were displayed at an online edit terminal.

C.4.5.7 Image-Processed Documents

The paper edit log for key-entered documents is replaced by online viewing of suspect data for all image-processed documents. For rapid resolution, the edit criteria for each item in question appeared on the screen along with the suspect item. Corrections were made immediately. The system employed an edit/verify system which ultimately meant that two different people viewed the same suspect data and operated on it separately. The verifier made sure the two responses (one from either the entry operator or the ICR engine) were the same before the system accepted that item as being correct. If the editor could not determine the appropriate response, he or she escalated the suspect situation to a supervisor. For errors or suspect information that could not be resolved by supervisory staff, a product-line queue was created. This allowed supervisors in the processing area to escalate edits to project staff for resolution.

Once an entire batch was through the edit phase, it became eligible for the count-verification phase. The administration schedule data were examined systematically for booklet IDs that should have been processed (assessed administration codes). All documents under that administration schedule were then inspected to ensure that all of the booklets were included.

With the satisfactory conclusion of the count-verification phase, the edited batch file was uploaded to the mainframe, where it went through yet another edit process. A paper edit log was produced and, if errors remained, was forwarded to another editor. When this edit was satisfied, the PCS and WFM tracking systems were updated.

C.4.5.8 OMR Scanned and Key-Entered Documents

All student documents on the administration schedule were accounted for, as receipt control personnel checked that the materials were undamaged and assembled correctly. The machine edits performed during data capture verified that each sheet of each document was present and that each field had an appropriate value. All batches entered into the system, whether key entered or machine scanned, were edited for errors.

Data editing took place after these checks. This consisted of a computerized edit review of each respondent's document and the clerical edits necessary to make corrections based upon the computer edit. This data editing step was repeated until all data were correct.

A computerized edit list, produced after NAEP documents were scanned or key entered, and all the supporting documentation sent from the field were used to perform the first phase of the edit function. The hard-copy edit list contained all the vital statistics about the batch: number of students, school code, type of document, assessment code, suspect cases, and record serial numbers. Using this information, the data editor verified that the batch had been assembled correctly and that each school number was correct. During data entry, counts of processed documents were generated by type. These counts were compared against the information captured from the administration schedules. The number of assessed and absent students processed had to match the numbers indicated on the PCS.

In the second phase of data editing, experienced editing staff used a predetermined set of specifications to review the field errors and record necessary corrections to the student data file. The computerized edit list used in phase one was used to perform this function. The editing staff reviewed the computer-generated edit log and the area of the source document that was noted as being suspect or as containing possible errors. The composition of the field was shown in the edit box. The editing staff checked this piece of information against the NAEP source document. At that point, one of the following took place:

- (a) Correctable error — If the error was correctable by the editing staff according to the editing specifications, the correction was noted on the edit log for later correction via key entry.
- (b) Alert — If an error was not correctable according to the specifications, an alert was issued to NAEP project staff for resolution. Once the correction information was obtained, the correction was noted on the edit log for key entry correction.
- (c) Noncorrectable error — If a suspected error was found to be correct as stated and no alteration was possible according to the source document and specifications, no corrective action was taken. The programs were tailored to allow this information to be accepted into the data record.

The corrected edit log was then forwarded to the key-entry staff for processing. When all corrections were entered and verified for a batch, an extract program pulled the corrected records into a mainframe dataset. At this point, the mainframe edit program was initiated. The edit criteria were again applied to all records. If there were further errors, a new edit listing was printed and the cycle was repeated.

When the edit process produced an error-free file, the booklet ID number was posted to the NAEP tracking file by age, assessment, and school. This permitted Pearson Educational Measurement staff to monitor the NAEP processing effort by accurately measuring the number of documents processed by form. The posting of booklet IDs also ensured that a booklet ID was not processed more than once.

C.4.6 Image Quality Assurance Items

To provide another quality check on the image scanning and scoring system, Pearson Educational Measurement staff implemented a quality check process by creating a stamp with a valid score designated on it.

Each unique item type scored via the image system had two quality assurance (QA) stamps per valid score. These stamps and mock scores were placed in blank unused booklets by clerical staff and were sent through the same scanning and editing process as all assessed student documents.

The QA booklets were batched and processed together with student documents of the same type. Since all of a specific item were batched together for transmission to the scoring facility, the QA-stamped responses were integrated with the student responses and transmitted simultaneously to the scoring facility. During the scoring process, both student responses and the QA items were randomly displayed so scores could be applied.

When a reader later saw the QA sample on the monitor during scoring, he or she was to notify the table leader, who confirmed the score assigned by the reader was the score listed on the sample.

All image quality-assurance documents were created prior to the beginning of scoring and all pre-determined score points were used. During the process of scoring, valid score points could be changed or dropped due to revisions in the scoring rubrics. Pearson Educational Measurement provided ETS with documentation as to what score points on these items are no longer valid. When an image QA stamp was displayed to a reader that contained a score point that was no longer valid, the reader gave the response a score point of OT.

The Pearson Educational Measurement/NAEP PCS system produced various status reports. The Receipt Control Status Report was designed to track the receipt of material from the schools. It was sorted by school number and displayed the following information: Participation Status, Scheduled Administration Date, and the Shipment Receipt Date. The comment field in this report showed any school for which a shipment had not been received within three days of the scheduled test date.

The Processing Status Report was divided into two sections. The first section was sorted by school and grade within each assessment. The following preliminary data for each were entered off of the administration schedule as the shipment was opened by the receiving department:

- school number,
- session code,
- test date,
- preliminary count date,
- preliminary to be assessed,
- preliminary total assessed, and
- completeness flag (if appropriate.)

Preliminary Count information was updated programatically with the following Actual count data at the completion of processing:

- actual to be assessed ,
- actual total assessed,
- actual number withdrawn and ineligible,
- actual count date,
- actual number excluded, and
- actual number absent.

Section two of the report, called the PSC Exceptions Report, was newly developed by Pearson Educational Measurement for this processing cycle with input from ETS and Westat staff. Designed for

quick reference during the assessment period, the Exceptions Report listed only those schools and sessions with a discrepancy in processing (i.e., sessions not returned within three days of the scheduled assessment date or schools/sessions given a completeness flag). Once all discrepancies were resolved for a school, the school would be taken off of the Exceptions Report.

An Alert Report listed all information and problem alerts recorded on the online system in the PCS for a school or session.

Pearson Educational Measurement transmitted an electronic file to Westat weekly which contained the above data. Hard copies of the PCS Exceptions, Alerts, and Documents Processed Reports were also sent to ETS and Westat weekly.

C.4.7 Data Transmission Before Scoring

Delivery of data to the scoring center was accomplished via T1 transmission lines linking the mainframe computers and the NAEP servers at the document scanning site in the Pearson Educational Measurement main facility with the scoring servers dedicated to distributing work to the professional readers at the scoring center. The actual task of scheduling items for downloading was accomplished using a code written by the Image Software Development team. This code enabled the person scheduling the download to choose a team of readers and select the scheduled items from a list of all items which that team would be scoring throughout the scoring project. This process was repeated for all teams of readers until all anticipated work was scheduled. Once this task was completed, the scheduled job was tested to determine if there was sufficient free disk space on the servers at the scoring center. If for any reason sufficient disk space was not available, scheduled items could be deleted from the batch individually or as a group until the scheduled batch job could accommodate all items on the available disk space at the scoring center. Once it was determined that sufficient disk space was available, transmission of student responses commenced. Data transmission was typically accomplished during off-shift hours to minimize the impact on system-load capacity.

C.5 SCORING OVERVIEW

The NAEP 1997 Arts Assessment was administered at grade 8 in the disciplines of Music, Theatre, and Visual Arts. The grade 12 field test was administered in the disciplines of Music, Theatre, Visual Arts, *and* Dance. Each discipline included cognitive items (A/B blocks, most of which were scored on the Image system) and *Performing*-based items (C blocks). A variety of item types were used, including both written responses and *Performing* exercises that asked students to create their own works of art. Because of the specialized nature of the assessment components, the Pearson Educational Measurement Performance Assessment Scoring Center hired highly qualified individuals to review student materials as they became available, prepare training materials from sample responses and *Performings*, and lead teams through the scoring process. Also, because of the complexity of the project design and the specialized nature of each assessment block, Pearson Educational Measurement worked with ETS trainers to conduct rangefinding only after most of the materials were through processing. Training and scoring took place only after all materials were at Pearson Educational Measurement. With the proper training materials and equipment in place, Pearson Educational Measurement proceeded to score the NAEP 1997 Arts Assessment and grade 12 field test.

C.5.1 Arts Image Test Overview

The 1997 Assessment and field test used a variety of constructed-response items to measure different elements of students' knowledge, ability to interpret, and creativity in the arts. These items were administered in scannable assessment booklets. The items scored included dichotomous items, short constructed responses, extended constructed responses, diagrams, and drawings. Each item had a unique scoring guide that identified the range of possible scores for the item and defined the criteria to be used in evaluating student responses.

The readers scoring the constructed-response items were organized into teams. Each team scored blocks from both the eighth and twelfth grades. An attempt was made to assign discrete blocks to teams, but occasionally individual items were divided among teams in order to better utilize scorers' particular areas of expertise. All responses were scored on scales ranging from 2 to 5 points.

C.5.2 Arts Performance Test Overview

Certain C-block booklets involved short-answer written responses, some objective and some self-reflective, and actual student *Performings* that were scored for a variety of relevant qualities. All responses were scored on scales ranging from 2 to 5 points. Theatre and Dance *Performings* were recorded on video tape. Music *Performings* were recorded on audio tape. Visual arts production work was submitted in a number of media including clay (scored using Polaroid photos instead of original work), pencil, charcoal, and pen.

The readers scoring the *Performing* items were organized into teams. Teams scored blocks from both the eighth and twelfth grades. Individual blocks were generally scored by a single team. However, in a few cases they were divided between teams to make best use of scorers' specialized expertise. Scoring scales were created specifically for each individual item. Any one creative piece a student produced, for example, might potentially have been scored on a number of different scales for each of several qualities or aspects of the work.

C.5.3 Arts Staffing and Schedule

One *Performing* assessment specialist and one team leader over-saw all four disciplines. Because of the scale of the arts portion of NAEP, as well as the need for particular expertise in those overseeing the project, Pearson Educational Measurement staffing of the project was slightly more specialized than for typical projects. Every effort was made to hire scorers with degrees and/or professional experience in the discipline in which they were scoring. Many scorers had teaching experience. All scorers hired had course work and/or practical experience in the content field.

The assignment of readers to the teams was made on the basis of each reader's academic and practical experience in fine arts with an emphasis upon whichever medium or style was being scored, as well as on other scoring-related experience.

Scoring-related activities for the four arts disciplines occurred on a staggered timetable, which generally involved two disciplines working simultaneously. Please refer to Table C-2 for specific time frames for each.

C.5.4 Arts Training Exemplar Selection

For both Image and Performance items, the 1997 arts training exemplar selection began three to five weeks prior to scoring with Pearson Educational Measurement table leaders and ETS trainers reading a sampling of student responses and making decisions about how to apply scoring guides. Although the selection process for each of the four arts disciplines varied slightly, it was generally carried out as illustrated in the following two sections.

Table C-2
NAEP 1997 Arts Assessment and Grade 12 Arts Field Test Dates

Assessment	Number of Table Leaders	Number of Scorers	Dates
Dance A/B	1	9	06/02/97—06/11/97
Dance Performance	1	11	06/02/97—06/14/97
Music A/B	2	19	06/23/97—07/17/97
Music Performance	5	48	06/23/97—07/28/97
Theatre A/B	2	7	06/17/97—07/09/97
Theatre Performance	4	29	06/17/97—07/19/97
Visual Arts Part 1	5	44	06/02/97—06/24/97
Visual Arts Part 2	2	9	07/07/97—07/15/97

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

C.5.4.1 Arts Image Paper Selection

After reading student responses, table leaders and trainers began assigning tentative score levels, as well as working to find as much of a range of levels of student *Performing* as possible. Table leaders and trainers usually had time to compile Anchor Sets for each item during the one or two weeks that trainers were present. After ETS trainers completed the initial round of paper selection, table leaders continued the process by compiling two practice sets for each item.

For grade 8 items, two qualifying sets were also compiled for each item. Anchor sets were usually comprised of two or three clear examples per score point. Practice sets included 10 to 25 prescored practice papers that represented the entire range of score points the item could receive. Each qualifying set included a total of six to seven examples for 3-point items and ten examples for 4- and 5-point items.

C.5.4.2 Arts Performing Paper Selection

The selection process for *Performing* blocks was the same as for Image; however, for Music, Theatre, and Dance, table leaders dubbed training tapes as well. Visual arts table leaders prepared training exemplars that were actual student works in such a way that they could be “exhibited” to scorers during training.

C.5.5 Arts Scorer Training

The training for the fine arts was conducted by ETS trainers/arts specialists, with assistance from Pearson Educational Measurement table leaders. The following two sections detail how training proceeded for Image and for Performance blocks.

C.5.5.1 Image Training

Training involved explaining the item and its scoring guide to the team and discussing responses that represented the various score points in the guide. Typically two or three anchor responses were chosen for each score point. During this stage readers and the table leaders kept notes of scoring decisions. The table leaders were then responsible for compiling those notes and ensuring that all readers were in alignment. When review of the anchor packet was completed, the readers scored and discussed 10 to 25 prescored practice papers that represented the entire range of score points the item could receive.

For the Arts Assessment and the grade 12 field test, the teams worked through qualification sets on paper for grade 8 only. Qualification sets generally consisted of 6–10 papers of which scorers were required to score 60 percent in agreement with the trainer and table leader. Scorers worked on grade 8 items only after successful completion of the qualification sets. For grade 12 items, once training was complete and the trainer and table leader judged the team to be ready to score, the table leader routed work to all scorers, and scoring commenced. In either case, to make sure that all scorers had a common understanding of the training, the teams usually began by group scoring several live student responses. Then the teams broke into pairs for scoring, followed by individual scoring when the ETS trainer was confident that all understood and could implement the scoring guidelines.

When scoring on the Image system, each team was trained to score a single item at a time. However, items that had multiple guides or were closely related were trained and scored together as “linked” items. On the Image screen, a red box indicated which component was being scored. As the readers scored one aspect of the response, the red box automatically jumped to the next aspect to be scored. With Theatre and Visual Arts, some cognitive items were scored on paper because their design did not allow Image scoring. Each team typically scored one to three items per day.

C.5.5.2 Performing Training

Because of the number of items to be scored, related items were clustered for training and scoring. Training procedures for *Performing* items were the same as for Image. For blocks where students were producing actual pieces of art, these had to be shown simultaneously on video screens for Theatre and Dance, played on audio tape for Music, and displayed in an exhibit fashion around the walls of the scoring area for Visual Arts. Written responses were trained using photocopied training sets, although some of these responses called for students to reflect on their *Performings*, and so had to be read while also viewing the actual *Performing*.

When anchor sets were completed, the readers scored and discussed practice sets of exemplars that included both easy examples and more difficult examples for all score points. Readers were given anchor sets for individual items in a cluster. Practice sets were usually clustered for training, but were sometimes organized by individual item. Qualification requirements and procedures were the same for grade 8 *Performing* items as for Image training, with the exception that students viewed or listened to *Performings* as part of their qualification sets.

C.5.6 Arts Scoring

Once the practice sessions were completed for both Image and Performance, the formal scoring processes began. The following two sections describe in detail the procedures that were used for the scoring of the 1997 arts for the various disciplines. For all arts disciplines, information regarding total number of Image and *Performing* items scored may be found in Table C-3.

Table C-3
NAEP 1997 Grade 8 Arts Assessment
Number of Constructed-Response Items

Component	Polytomously Scored Items				Total
	Dichotomous 2-point	3-point items	4-point items	5-point items	
Grade 8 Music Items					
<i>AB-Image</i>	12	41	7	1	61
<i>C-Performing</i>	20	9	37	11	77
Grade 8 Theatre Items					
<i>AB-Image</i>	0	17	11	0	28
<i>AB-Paper</i>	0	0	2	0	2
<i>C-Performing</i>	0	9	13	0	22
<i>C-Evaluations</i>	1	9	1	0	11
Grade 8 Visual Arts Items					
<i>AB-Image</i>	3	11	4	0	18
<i>AB-Paper</i>	2	3	6	2	13
<i>C-Paper</i>	0	13	13	1	27

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

C.5.6.1 Image Scoring

During Image scoring, the table leaders continued to compile notes on scoring decisions for the readers' reference and guidance. Additionally, the table leaders closely monitored interreader reliability using both team and individual statistics as a reference. Table leaders also performed backreading duties in which they reviewed a sample of the responses scored by each reader on the team. The team leader and *Performing* assessment specialist continuously monitored the progress of each team to ensure that training and scoring progressed smoothly and in a timely manner.

The backreading software tool allowed Pearson Educational Measurement table leaders to monitor each reader's progress by reviewing papers scored by each reader on the team. Typically, a table leader reviewed approximately 10 percent of all responses scored by each reader. Table leaders made certain to note the score the reader awarded each response as well as the score a second reader gave that same paper. This was done as an interreader reliability check. Alternatively, a table leader could choose to review all responses given a particular score to determine if the team as a whole was scoring consistently. Both of these review methods used the same display screen and showed the ID number of the reader and the scores awarded. If the table leader disagreed with the score given an item, he or she discussed it with the reader for possible correction. Replacement of scores by the table leader was done only with the knowledge and approval of the reader, thereby serving as a learning experience for the

reader. Scores were changed only if the first score was incorrect, in order to get the most accurate scores for analysis. The system was modified this year so that changing the score did not change the measurement of interreader reliability.

The Theatre assessment used a variety of constructed-response items in order to measure different elements of students' theatre knowledge, interpretation ability, and creativity in script writing, stage craft, costuming, set design, and other aspects of theatre. Two of the items were scored on paper. The various Music items measured different elements of students' musical sophistication, interpretation, and comprehension. Items were assigned to the teams according to the complexity and number of items within each block. The Visual Arts items were designed to measure students' visual-arts knowledge, interpretation capabilities, and creativity. The readers scoring the constructed-response items were divided into two different teams. Each of these teams scored both Image items and *Performing*-based items wherein they rated actual student artwork. One team scored the Dance portion of the 1997 field test. A variety of constructed-response items were used to measure different elements of students' knowledge and ability to interpret Dance *Performings*.

C.5.6.2 Performing Scoring

Clerical support staff circulated boxes of student works (video tapes for Theatre and Dance; audio tapes for Music; student production work for Visual Arts) to appropriate teams for scoring. Prior to scoring, arts materials were sorted by batch, and labeled with batch numbers, block ID's and student ID's. Score sheets were sorted by individual cluster and matched with clusters of student responses. Additionally, score sheets for Theatre and Dance were ordered by student ID and matched with tapes.

Readers for Music, Theatre, and Dance used headphones while listening to the tapes to achieve optimal sound quality without disturbing neighboring scorers. When items being scored were written responses, student booklets were distributed for scoring. During the scoring, the table leader compiled notes on various items for the readers' reference and guidance. In addition, table leaders rigorously conducted backreading of the scorers, which entailed observing individuals as they scored and comparing the scores readers assigned to the scores the table leaders themselves would assign. Any readers who were noted still to be having problems internalizing the scoring guides were provided additional training. When particular items seemed to give the majority of the group members problems, trainers would halt the scoring for brief retraining sessions.

Each block for arts *Performing* included one or occasionally two performing activities in which students participated. These exercises were designed to assess students' technical abilities and experience in the arts. Each discipline sought to encompass as broad a range as possible. In Theatre, students both interpreted existing scripts and performed improvised scenes. For Music, students were asked to compose and sight-read music, and learn rhythms and melodies by ear. They were also given the opportunity to improvise in Jazz and Rock styles. In Visual Arts, students worked with a variety of media using paintings, sculptures, stories, and other stimulus materials to inspire their own creative works. In Dance, students were taught a choreographed Jazz piece, and were given the opportunity to choreograph their own piece based on a literary stimulus. The student *Performings* were scored on various scales that incorporated qualities crucial to the particular disciplines, such as use of body and voice for Theatre, or use of color and texture for Visual Arts. For some blocks, students were also asked to write short constructed responses that usually asked them to reflect upon the process of creating their work.

Scores were recorded onto scannable scoring sheets. For Music, Theatre, and Dance, each rater used a different sheet for each cluster of guides for each student *Performing*. For Visual Arts, in order to limit the number of score sheets, every student in a batch was listed on one score sheet per item. These

processes allowed for scanning into the database and made feedback available on the level of the individual scorers. Unless otherwise noted, all scoring was done by individual raters.

Panel scoring: In an effort to reflect a more realistic situation for scoring or judging works of student art, panel scoring was introduced into the scoring model for Theatre and Dance. A panel was a 2- or 3-member subgroup of a team that viewed a student work simultaneously on the same monitor with each person listening on his or her own headphones. The scorers listened to then discussed the *Performing*, recording their individual scores. Panel members were encouraged to come to agreement on their interpretation of the scoring guide, but were not required to give the same score to a student, although generally they did. At the end of Theatre scoring, due to time constraints, some twelfth grade Theatre *Performing* items were scored individually rather than by panels.

C.5.7 Reliability of Scoring

Twenty-five percent of all Music and Visual Arts responses for each item or student work was scored by a second reader for all Image and *Performing* items. For Theatre and Dance, twenty-five percent of all Image items were second-scored; one-hundred percent of the *Performing* items were second-scored. The general goal for reliability was 70 percent agreement or greater for 4- or 5-point items, and 80 percent or greater for items with 3 or fewer score points. Interreader reliability ranges can be found in Table C-4. This reliability information was also used by the table leaders in monitoring the capabilities of all readers and uniformity of scoring across readers. The following two sections briefly describe how reliability was gauged in the arts field test for both Image and Performance scoring.

Table C-4
NAEP 1997 Grade 8 Arts Assessment
Interreader Reliability Ranges

Assessment	Number of Unique Items	Number of Items in Percentage Agreement Range				
		Below 60%	60—69%	70—79%	80—89%	Above 90%
Music AB- Image	61	0	0	3	21	37
Music C-Performance	77	0	6	24	26	21
Theatre AB-Image	28	0	0	7	15	6
Theatre AB-Paper	2	0	0	0	1	1
Theatre C-Performance	22	0	0	4	11	7
Theatre C-Evaluation	11	0	0	0	5	6
Visual Arts AB-Image	18	0	0	6	11	1
Visual Arts AB-Paper	13	0	0	0	3	10
Visual Arts C-Paper	27	0	0	6	16	5

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

C.5.7.1 Rescore

To determine how reliably the readers were scoring responses to specific items, in cases where reliability rates were not high enough according to professional judgment, ETS personnel, in consultation with Pearson Educational Measurement staff, made the decision to rescore items. Items that were

rescored are T2MU9, items 1A and 1C; T2TH6_01; T2TH9, item Group 1C; T2VA5_03 and 04; and T3VAX10, item 1.

C.5.7.2 Image Reliability

Reliability reports for Image scoring could be generated on demand by the table leaders when needed. Reports were available at both the team and scorer level. Although feedback was nearly instantaneous, a table leader could also review the actual responses scored by a reader by using the backreading tool. In this way the table leader was able to monitor each reader carefully and correct any difficulties in scoring almost immediately.

C.5.7.3 Performing Reliability

As scoring on a cluster of items progressed, the score sheets were scanned into the Pearson scoring system. Only sheets marked with valid ID numbers and one, and only one, valid score per item were accepted by the scanning system. Interreader and inter-panel reliability reports could be generated for the team and the individual. Reliability reports were generated twice daily when possible, and were used by table leaders and trainers to monitor readers and indicate which readers, panels, or teams needed retraining. For items that received scores by individual raters, agreement refers to whether the score assigned by Reader One is an exact match with the score assigned by Reader Two. For panel scoring, agreement refers to an exact match of the sum of the three scores assigned by the members of Panel One with the sum of the three scores assigned by the members of Panel Two.

C.6 PREPARATION FOR TAPE CREATION

The Arts Assessment and the grade 12 field test data collection resulted in several classes of data files—student, school, teacher, SD/LEP student, student/teacher match, and student-response information. Student-response information included response data from all assessed students in 1997. Data resolution activities occurred prior to the submission of data files to ETS and Westat to resolve any irregularities that existed. This section details additional steps performed before creating the final data files to ensure capture of the most complete and accurate information.

C.7 UPLOADING OF SCORES TO DATABASE

An important quality-control component of the image-scoring system was the inclusion, for purposes of file identification, of an exact copy of the student edit record, including the student booklet ID number, with every image of a student's response to a constructed-response item. These edit files also remained in the main data files residing on the Pearson Educational Measurement mainframe computer. By doing this, exact matching of scores assigned to constructed-response items and all other data for each individual student was guaranteed, since the booklet ID for each image was part of every image file.

When all the responses for an individual item had been scored, the system automatically submitted all item scores assigned during scoring, along with their edit records, to a queue to be transmitted to the mainframe. Development staff then initiated a system job to transmit all scoring data to be matched with the original student records on the mainframe. A custom edit program matched the edit records of the scoring files to those of the original edit records on the mainframe. As matches were confirmed, the scores were applied to those individual files. After completion of this stage, all data

collected for an individual student was located in one single and complete record/file identified by the edit record.

C.8 SD/LEP QUESTIONNAIRES

Pearson Educational Measurement processed the SD/LEP questionnaires via OMR scanning. Edits performed on the questionnaires assured that responses to questions fell within the valid range for that question. SD/LEP questionnaires were then matched to a student record. SD/LEP questionnaires that were not matched to a student document were cross-referenced with the corresponding Administration Schedule, Roster of Questionnaire, and student data files to correct, if necessary, the information needed to result in a match.

C.9 SCHOOL QUESTIONNAIRES

In 1997, Pearson Educational Measurement continued to use ICR technology to capture percentage figures written by school personnel directly in boxes on the school questionnaire rather than requiring the school official to grid ovals in a matrix. The data was then verified by an edit operator.

C.10 TEACHER QUESTIONNAIRE MATCH

The same processes that were followed in previous cycles were used in 1997 to achieve the best possible student/teacher match rate. Student identification numbers that were not matched to a teacher questionnaire were cross-referenced with the corresponding Administration Schedule and Roster of Questionnaires to verify (and change, if necessary) the teacher number, teacher period, and questionnaire number recorded on these control documents. The NAEP school numbers listed on the Roster of Questionnaires and teacher questionnaire were verified and corrected, if necessary. Once these changes were made, any duplicate teacher numbers existing within a school were, if possible, cross-referenced for resolution with the Rosters of Questionnaires. Since this information was located together on a single, central control document, the ability to match and resolve discrepant or missing fields was simplified.

C.11 DELIVERY AND STORAGE OF DOCUMENTS

After all data-processing activities were completed, data cartridges and/or diskettes were created and shipped via overnight delivery to ETS and/or Westat. Pearson Educational Measurement maintains a duplicate archive file for security/back-up purposes.

After batches of documents had successfully passed the scoring process, they were sent to the Pearson Educational Measurement warehouse for storage. Due to the number of items related to a single student the A/B book, C-block book, and related student artwork were placed into permanent storage by their processing batch number. This will allow for efficient document and artwork retrieval to fill requests for specific student information in the future. Once documents and associated artwork were boxed by batch, each grade and subject area were assigned a unique inventory number for the boxes to be stored under. The storage locations of all documents were recorded on the inventory control system.

Unused materials were sent to temporary storage to await completion of the entire assessment. Once the assessment was complete, Pearson Educational Measurement received authorization from ETS to salvage unused materials after determining that a sufficient quantity of each form type was retained in permanent archive.

C.12 QUALITY-CONTROL DOCUMENTS

ETS required that a random sample of booklets and score sheets be pulled for an additional quality-control check. For image scored items, a scoring sheet is not used, so ETS uses scores sent to them on a data tape to verify the accuracy of applied scores. All of the above documents were selected prior to sending the booklets and score sheets to storage and were sent to ETS to verify the accuracy and completeness of the data. A random sample of all the questionnaires used in the NAEP 1997 Arts Assessment were also sent to ETS along with the Quality Assurance booklets used for processing and scoring.

Appendix D

NAEP 1997 ARTS ASSESSMENT SCORING RELIABILITY

This appendix contains information about the constructed-response items in the 1997 Arts Assessment of Music, Theatre, and Visual Arts. For each of these three arts disciplines, the information in the tables includes the NAEP item numbers for each of the constructed-response items included in scaling, and the block that contains the item. The tables also indicate the codes from the NAEP database that denote the range of responses. A portion of the responses to the constructed-response items were scored twice for the purpose of examining rater reliability. For each item, the number of papers with responses that were scored a second time is listed, along with the percent agreement between raters and an index of reliability based on those responses. Cohen's Kappa (Cohen, 1968) is the reliability estimate used for dichotomously scored items. For polytomous items, the intraclass correlation coefficient is used as the index of reliability.

Table D-1
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Cohen's Kappa¹
for the Dichotomously Scored Constructed-Response Music Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
UC000A2	MC	†	1 – 2	299	97	0.94
UC000B2	MC	†	1 – 2	299	98	0.95
UC00004	MC	Responding 1	1 – 2	299	100	0.99
UC00012	MC	Responding 1	1 – 2	299	99	0.99
UD000A5	MD	Responding 2	1 – 2	277	100	0.99
UD000B5	MD	Responding 2	1 – 2	277	99	0.98
UD000C5	MD	Responding 2	1 – 2	277	90	0.82
UD000D5	MD	†	1 – 2	275	96	0.93
UD000C6	MD	Responding 2	1 – 2	275	96	0.92
UE000A6	ME	†	1 – 2	279	99	0.99
UE000B6	ME	†	1 – 2	279	100	1.00
UE000C8	ME	Creating	1 – 2	279	98	0.97
UH00004	MH	†	1 – 2	155	99	0.93
UI000D1	MI	Creating	1 – 2	165	95	0.87
UJ000A3	MJ	†	1 – 2	63	97	0.94
UJ000A4	MJ	†	1 – 2	63	98	0.97
UJ000A5	MJ	†	1 – 2	63	95	0.91
UJ000A6	MJ	†	1 – 2	63	98	0.97
UJ000A7	MJ	†	1 – 2	62	97	0.94
UJ000A8	MJ	†	1 – 2	62	97	0.94
UJ000A9	MJ	†	1 – 2	61	98	0.97
UJ000A0	MJ	†	1 – 2	61	93	0.88
UJ000B3	MJ	†	1 – 2	63	90	0.82
UJ000B4	MJ	†	1 – 2	63	97	0.91
UJ000B5	MJ	†	1 – 2	63	84	0.70
UJ000B6	MJ	†	1 – 2	63	92	0.85
UJ000B7	MJ	†	1 – 2	62	95	0.91
UJ000B8	MJ	†	1 – 2	62	97	0.92
UJ000B9	MJ	†	1 – 2	61	97	0.91
UJ000B0	MJ	†	1 – 2	61	87	0.72
UJ000D3	MJ	†	1 – 2	63	94	0.88
UJ000D4	MJ	†	1 – 2	61	92	0.81

¹ Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomously scored. These items are scored right or wrong.

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table D-2
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Music Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
UC00006	MC	Responding 1	1 - 3	299	91	0.93
UC00010	MC	Responding 1	1 - 4	299	96	0.98
UC000A3	MC	†	1 - 3	299	92	0.93
UC000B1	MC	†	1 - 3	299	91	0.86
UC000B3	MC	†	1 - 3	299	89	0.90
UC000B14	MC	†	1 - 3	298	93	0.90
UC000B15	MC	†	1 - 3	288	92	0.90
UC000C14	MC	†	1 - 3	298	89	0.88
UC000C15	MC	†	1 - 3	283	89	0.81
UC000D15	MC	†	1 - 3	259	90	0.83
UD00007	MD	Responding 2	1 - 3	275	85	0.88
UD000A6	MD	Responding 2	1 - 3	275	95	0.97
UD000B6	MD	Responding 2	1 - 3	275	89	0.90
UD000B9	MD	†	1 - 3	275	99	0.98
UD000D6	MD	†	1 - 3	275	97	0.97
UD000E5	MD	†	1 - 3	275	95	0.94
UD000E6	MD	†	1 - 3	275	97	0.96
UE00004	ME	Responding 1	1 - 4	279	99	1.00
UE00010	ME	Responding 2	1 - 3	279	84	0.85
UE00015	ME	Responding 2	1 - 3	274	84	0.72
UE00016	ME	Responding 2	1 - 3	263	86	0.89
UE000A8	ME	†	1 - 3	279	94	0.96
UE000A9	ME	†	1 - 3	279	93	0.95
UE000B8	ME	†	1 - 3	279	96	0.96
UE000B9	ME	†	1 - 3	279	90	0.94
UE000C9	ME	†	1 - 3	279	94	0.98
UE000B11	ME	†	1 - 3	279	94	0.95
UE000B12	ME	†	1 - 3	279	98	0.99
UF00003	MF	Responding 2	1 - 4	278	92	0.96
UF00005	MF	Responding 2	1 - 5	277	97	0.93
UF000B1	MF	†	1 - 3	279	95	0.97
UF000B2	MF	†	1 - 3	279	99	0.98
UF000H1	MF	†	1 - 3	279	96	0.98
UF000L1	MF	†	1 - 3	279	95	0.98
UF000N1	MF	†	1 - 3	279	97	0.98
UF000P1	MF	†	1 - 3	279	98	0.98
UF000T1	MF	†	1 - 3	279	96	0.98

(continued)

Table D-2 (continued)
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Music Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
UF001A6	MF	†	1 – 3	276	97	0.96
UF001B6	MF	†	1 – 4	276	92	0.94
UF001C6	MF	†	1 – 3	276	96	0.96
UF002A6	MF	†	1 – 3	276	99	0.99
UF002B6	MF	†	1 – 4	276	93	0.92
UF002C6	MF	†	1 – 3	276	93	0.92
UF003A6	MF	†	1 – 3	276	99	0.99
UF003B6	MF	†	1 – 4	275	96	0.97
UF003C6	MF	†	1 – 3	275	93	0.92
UF004A6	MF	†	1 – 3	275	100	0.96
UF004B6	MF	†	1 – 4	272	99	0.99
UF004C6	MF	†	1 – 3	266	98	0.99
UG000A1	MG	Creating	1 – 4	164	78	0.90
UG000A2	MG	Performing	1 – 4	160	91	0.94
UG000B1	MG	Creating	1 – 4	164	81	0.87
UG000B2	MG	Performing	1 – 4	160	80	0.84
UG000C1	MG	Creating	1 – 4	159	66	0.76
UH000A1	MH	Performing	1 – 4	149	91	0.95
UH000A2	MH	Performing	1 – 3	154	81	0.89
UH000B1	MH	Performing	1 – 4	149	85	0.94
UH000B2	MH	Performing	1 – 4	154	76	0.92
UH000C1	MH	Creating	1 – 4	155	79	0.87
UH000C2	MH	Performing	1 – 4	154	68	0.81
UH000D2	MH	Performing	1 – 4	154	60	0.83
UH000E2	MH	Performing	1 – 3	154	70	0.80
UI00002	MI	†	1 – 3	166	79	0.69
UI00003	MI	†	1 – 3	161	86	0.79
UI000A1	MI	Creating	1 – 4	167	75	0.85
UI000A3	MI	Creating	1 – 4	166	81	0.95
UI000B1	MI	Creating	1 – 4	166	70	0.83
UI000C1	MI	Creating	1 – 4	162	65	0.75
UI000C3	MI	Creating	1 – 4	166	87	0.89
UI000E1	MI	Creating	1 – 3	168	77	0.66
UJ000A1	MJ	†	1 – 4	55	73	0.85
UJ000A2	MJ	†	1 – 4	67	79	0.82
UJ000B2	MJ	†	1 – 4	67	73	0.85
UJ000C2	MJ	†	1 – 4	67	82	0.82

(continued)

Table D-2 (continued)
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Music Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
UJ000C3	MJ	†	1 - 4	61	97	0.86
UJ000D1	MJ	†	1 - 5	52	87	0.98
UJ000D2	MJ	†	1 - 4	67	87	0.88
UJ000E1	MJ	†	1 - 4	43	65	0.89
UJ000E2	MJ	†	1 - 4	67	84	0.82
UJ000F1	MJ	†	1 - 5	34	59	0.81
UJ000G1	MJ	†	1 - 5	52	73	0.93
UJ000H1	MJ	†	1 - 5	43	84	0.90
UJ000I1	MJ	†	1 - 4	68	76	0.88
UJ001A2	MJ	†	1 - 4	62	73	0.70
UJ001B2	MJ	†	1 - 4	62	76	0.79
UJ001C2	MJ	†	1 - 4	62	76	0.76
UJ001D2	MJ	†	1 - 4	62	84	0.83
UJ001E2	MJ	†	1 - 4	62	79	0.76
UK00003	MK	†	1 - 3	68	68	0.66
UK00004	MK	†	1 - 3	69	81	0.73
UK00005	MK	†	1 - 3	60	85	0.91
UK00006	MK	†	1 - 3	69	81	0.87
UK000A1	MK	†	1 - 4	56	73	0.88
UK000A2	MK	†	1 - 4	57	65	0.78
UK000B2	MK	†	1 - 5	72	86	0.94
UK000C2	MK	†	1 - 4	47	55	0.75
UK000D1	MK	†	1 - 5	48	79	0.96
UK000D2	MK	†	1 - 5	73	75	0.88
UK000E1	MK	†	1 - 4	47	74	0.87
UK000E2	MK	†	1 - 4	73	63	0.61
UK000F1	MK	†	1 - 5	32	75	0.94
UK000F2	MK	†	1 - 3	73	75	0.74
UK000G1	MK	†	1 - 5	51	75	0.94
UK000H1	MK	†	1 - 5	41	85	0.94
UK000I1	MK	†	1 - 4	72	71	0.84
UK00002	MK	†	1 - 5	32	63	0.81

¹ The intraclass correlation is most appropriate for items with more than two categories. They are reliability estimates that correct for chance agreement.

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table D-3
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Cohen's Kappa¹
for the Dichotomously Scored Constructed-Response Theatre Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
HG00003	TG	†	1 - 2	432	91	0.82

¹ Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomously scored. These items are scored right or wrong.

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table D-4
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Theatre Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
HC00002	TC	Responding	1 - 3	159	92	0.91
HC00004	TC	Responding	1 - 3	158	85	0.90
HC000B5	TC	Creating/Performing	1 - 4	159	89	0.94
HC000C5	TC	Creating/Performing	1 - 4	158	97	0.97
HC00006	TC	Responding	1 - 4	158	89	0.94
HC00007	TC	Creating/Performing	1 - 4	158	92	0.93
HC00008	TC	Responding	1 - 3	158	93	0.91
HC00009	TC	Responding	1 - 3	158	86	0.81
HD00003	TD	†	1 - 4	169	83	0.85
HD00004	TD	Responding	1 - 3	169	96	0.93
HD00005	TD	Creating/Performing	1 - 4	170	84	0.93
HD00007	TD	Creating/Performing	1 - 4	167	84	0.85
HD00008	TD	Creating/Performing	1 - 3	163	93	0.96
HD00009	TD	Responding	1 - 3	144	86	0.87
HE00001	TE	Responding	1 - 3	174	88	0.81
HE00002	TE	Responding	1 - 3	173	83	0.77
HE00003	TE	Responding	1 - 3	173	86	0.83
HE00004	TE	Creating/Performing	1 - 4	171	79	0.74
HE00005	TE	Responding	1 - 3	167	83	0.74
HE00006	TE	Responding	1 - 3	163	80	0.73
HE00007	TE	Responding	1 - 3	156	83	0.79
HF000B1	TF	Responding	1 - 4	189	79	0.77
HF00002	TF	Responding	1 - 4	188	80	0.87
HF00004	TF	Responding	1 - 3	188	95	0.95
HF00005	TF	Responding	1 - 4	188	82	0.75
HF00006	TF	Creating/Performing	1 - 4	188	77	0.84
HF00007	TF	Responding	1 - 4	188	95	0.86
HF00009	TF	Responding	1 - 3	186	89	0.77
HF00011	TF	Responding	1 - 3	185	82	0.77
HF00B12	TF	Responding	1 - 3	185	84	0.83
HG000A1	TG	Creating/Performing	1 - 4	431	96	0.87
HG000B1	TG	Creating/Performing	1 - 4	420	94	0.75
HG000C1	TG	Creating/Performing	1 - 4	427	84	0.81
HG000D1	TG	Creating/Performing	1 - 3	427	91	0.92
HG000E1	TG	Creating/Performing	1 - 3	427	87	0.82
HG000F1	TG	Creating/Performing	1 - 4	427	82	0.70
HG000G1	TG	†	1 - 4	431	78	0.74

(continued)

Table D-4 (continued)
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Theatre Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
HG00002	TG	†	1 - 3	432	91	0.56
HG00004	TG	†	1 - 3	429	89	0.83
HG00005	TG	†	1 - 3	425	88	0.84
HG00006	TG	†	1 - 3	426	88	0.83
HH000A1	TH	Creating/Performing	1 - 3	385	90	0.77
HH000B1	TH	Creating/Performing	1 - 3	385	87	0.87
HH000C1	TH	Creating/Performing	1 - 3	385	86	0.85
HH000D1	TH	Creating/Performing	1 - 3	385	86	0.86
HH000E1	TH	†	1 - 4	385	91	0.94
HH000F1	TH	Creating/Performing	1 - 4	385	82	0.89
HH000B2	TH	†	1 - 3	382	90	0.90
HH00003	TH	†	1 - 3	381	88	0.91
HI000A1	TI	Creating/Performing	1 - 4	390	97	0.96
HI000B1	TI	Creating/Performing	1 - 3	390	97	0.83
HI000C1	TI	Creating/Performing	1 - 3	390	84	0.83
HI000D1	TI	Creating/Performing	1 - 3	390	87	0.83
HI000E1	TI	Creating/Performing	1 - 3	390	97	0.76
HI000F1	TI	Creating/Performing	1 - 3	390	83	0.81
HI000G1	TI	†	1 - 4	388	87	0.90
HI000H1	TI	†	1 - 4	388	85	0.92
HI000I1	TI	†	1 - 4	388	93	0.84
HI00002	TI	†	1 - 3	390	92	0.85
HI00003	TI	†	1 - 4	387	93	0.96
HI00004	TI	†	1 - 3	362	91	0.92
HI00005	TI	†	1 - 3	359	91	0.93

¹ The intraclass correlation is most appropriate for items with more than two categories. They are reliability estimates that correct for chance agreement.

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table D-5
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Cohen's Kappa¹
for the Dichotomously Scored Constructed-Response Visual Arts Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Cohen's Kappa
VC000A6	VC	Creating	1 - 2	246	97	0.90
VD00001	VD	Responding	1 - 2	244	97	0.94
VD000B6	VD	Responding	1 - 2	243	88	0.76
VE00001	VE	Responding	1 - 2	238	87	0.71
VF000B7	VF	Creating	1 - 2	236	92	0.80

¹ Cohen's Kappa is a measure of reliability that is appropriate for items that are dichotomously scored. These items are scored right or wrong.

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Table D-6
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Visual Arts Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
VC00002	VC	Responding	1 - 3	245	88	0.91
VC00005	VC	Responding	1 - 4	243	91	0.92
VC00007	VC	Responding	1 - 4	194	92	0.95
VC000A4	VC	†	1 - 3	244	90	0.93
VC000B4	VC	†	1 - 3	244	75	0.83
VC000B6	VC	Creating	1 - 4	246	87	0.86
VD00007	VD	Responding	1 - 3	242	88	0.85
VD00008	VD	Responding	1 - 4	240	83	0.83
VD00009	VD	Responding	1 - 4	225	88	0.92
VD000A5	VD	Responding	1 - 3	243	88	0.86
VD000A6	VD	Responding	1 - 3	243	81	0.84
VD000B5	VD	Responding	1 - 3	243	82	0.85
VE00002	VE	Responding	1 - 4	238	82	0.75
VE00003	VE	Responding	1 - 4	238	80	0.82
VE00004	VE	Responding	1 - 3	238	80	0.82
VE00008	VE	Responding	1 - 3	238	86	0.91
VE00009	VE	Creating	1 - 5	238	91	0.92
VE00010	VE	Responding	1 - 3	185	94	0.81
VF00001	VF	Responding	1 - 3	235	83	0.81
VF00004	VF	Responding	1 - 3	235	82	0.80
VF000A6	VF	Responding	1 - 4	236	93	0.97
VF000A7	VF	Responding	1 - 3	236	94	0.96
VF000B6	VF	Creating	1 - 4	236	92	0.95
VF000B8	VF	Creating	1 - 4	236	89	0.93
VF000C6	VF	Creating	1 - 4	236	94	0.98
VF000C7	VF	Creating	1 - 3	236	91	0.93
VG00002	VG	†	1 - 3	151	89	0.90
VG000A1	VG	Creating	1 - 3	172	90	0.89
VG000B1	VG	Creating	1 - 3	172	84	0.79
VG000C1	VG	Creating	1 - 3	172	84	0.78
VG000D1	VG	Creating	1 - 3	172	86	0.73
VG000E1	VG	Creating	1 - 3	172	78	0.80
VG000F1	VG	Creating	1 - 3	172	85	0.82
VG000G1	VG	Creating	1 - 3	172	78	0.74
VH00001	VH	Creating	1 - 3	174	89	0.87
VH00004	VH	†	1 - 3	137	88	0.89
VH000A2	VH	Creating	1 - 4	174	84	0.83

(continued)

Table D-6 (continued)
NAEP 1997 Arts Assessment
Range of Response Codes, Percent Agreement, and Intraclass Correlation¹
for the Polytomously Scored Constructed-Response Visual Arts Items

Item	Block	Scale	Range of Response Codes	Sample Size	Percent Agreement	Intraclass Correlation
VH000A3	VH	Creating	1 - 4	174	80	0.84
VH000B2	VH	Creating	1 - 4	174	86	0.86
VH000B3	VH	Creating	1 - 3	174	80	0.85
VH000C2	VH	Creating	1 - 4	174	92	0.94
VH000C3	VH	Creating	1 - 4	174	83	0.88
VH000D3	VH	Creating	1 - 4	174	90	0.93
VI00001	VI	Creating	1 - 4	184	79	0.91
VI00002	VI	Creating	1 - 5	184	77	0.90
VI00004	VI	†	1 - 4	104	84	0.90
VI000A3	VI	Creating	1 - 4	181	81	0.87
VI000B3	VI	Creating	1 - 4	181	81	0.87
VI000C3	VI	Creating	1 - 4	181	82	0.88
VI000F3	VI	Creating	1 - 3	181	83	0.85
VI000G3	VI	Creating	1 - 3	181	87	0.78
VI000H3	VI	Creating	1 - 3	181	85	0.85
VI000I3	VI	Creating	1 - 4	181	80	0.84

¹ The intraclass correlation is most appropriate for items with more than two categories. They are reliability estimates that correct for chance agreement.

† Not applicable

SOURCE: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1997 Arts Assessment.

Appendix E

NAEP 1997 ARTS ASSESSMENT ITEMS CONTRIBUTING TO EACH ARTS SCALE

This appendix lists the items contributing to each of the three arts discipline scales (Music, Theatre, and Visual Arts) created for the NAEP 1997 Arts Assessment. The type of item is also provided:

- **Dichotomous item** — an item for which two possible responses (other than missing and off-task) exist: correct and incorrect; a common multiple-choice item or an item that requires a constructed response (noted as *Dichotomous C-R* within the following tables) from the student and is subsequently scored by scorers into two categories
- **Polytomous item** — an item for which more than two possible responses (other than missing and off-task) exist (noted as *Polytomous C-R* within the following tables)
- **Cluster items** — an item formed by collapsing two or more items following scoring

Table E-1
NAEP 1997 Arts Assessment
Items Contributing to the
Music Creating Scale

Item	Item Type
UE000C8	Dichotomous C-R
UECL008	Cluster
UG000A1	Polytomous C-R
UG000B1	Polytomous C-R
UG000C1	Polytomous C-R
UH000C1	Polytomous C-R
UI000A1	Polytomous C-R
UI000B1	Polytomous C-R
UI000C1	Polytomous C-R
UI000D1	Dichotomous C-R
UI000E1	Polytomous C-R
UI000A3	Polytomous C-R
UI000C3	Polytomous C-R

Table E-2
NAEP 1997 Arts Assessment
Items Contributing to the
Music Performing Scale

Item	Item Type
UG000A2	Polytomous C-R
UG000B2	Polytomous C-R
UH000A1	Polytomous C-R
UH000B1	Polytomous C-R
UH000A2	Polytomous C-R
UH000B2	Polytomous C-R
UH000C2	Polytomous C-R
UH000D2	Polytomous C-R
UH000E2	Polytomous C-R

Table E-3
NAEP 1997 Arts Assessment
Items Contributing to
Responding Scale 1

Item	Item Type
UCCL002	Cluster
UCCL003	Cluster
UC00004	Dichotomous C-R
UC00005	Multiple-choice
UC00006	Polytomous C-R
UC00007	Multiple-choice
UCCL008	Cluster
UC00009	Multiple-choice
UC00010	Polytomous C-R
UC00011	Multiple-choice
UC00012	Dichotomous C-R
UC00013	Multiple-choice
UCCL014	Cluster
UCCL015	Cluster
UD00004	Multiple-choice
UDCL005	Cluster
UDCL006	Cluster
UD00008	Multiple-choice
UD00010	Multiple-choice
UD00011	Multiple-choice
UE00004	Polytomous C-R
UE00005	Multiple-choice
UECL006	Cluster
UE00007	Multiple-choice
UFCL002	Cluster
UFCL0A6	Cluster
UFCL0B6	Cluster
UFCL0C6	Cluster
UFCL0D6	Cluster
UF00007	Multiple-choice
UF00008	Multiple-choice

Table E-4
NAEP 1997 Arts Assessment
Items Contributing to
Responding Scale 2

Item	Item Type
UCCL001	Cluster
UD00001	Multiple-choice
UD00002	Multiple-choice
UD00003	Multiple-choice
UD000A5	Dichotomous C-R
UD000B5	Dichotomous C-R
UD000C5	Dichotomous C-R
UD000A6	Polytomous C-R
UD000B6	Polytomous C-R
UD000C6	Dichotomous C-R
UD00007	Polytomous C-R
UDCL009	Cluster
UECL001	Cluster
UE00002	Multiple-choice
UE00003	Multiple-choice
UECL009	Cluster
UE00010	Polytomous C-R
UECL011	Cluster
UECL012	Cluster
UE00014	Multiple-choice
UE00015	Polytomous C-R
UE00016	Polytomous C-R
UF00003	Polytomous C-R
UF00004	Multiple-choice
UF00005	Polytomous C-R

Table E-5
NAEP 1997 Arts Assessment
Items Contributing to the
Theatre Creating/Performing Scale

Item	Item Type
HC000B5	Polytomous C-R
HC000C5	Polytomous C-R
HC00007	Polytomous C-R
HD00005	Polytomous C-R
HD00006	Multiple-choice
HD00007	Polytomous C-R
HD00008	Polytomous C-R
HE00004	Polytomous C-R
HF00006	Polytomous C-R
HG000A1	Polytomous C-R
HG000B1	Polytomous C-R
HG000C1	Polytomous C-R
HG000D1	Polytomous C-R
HG000E1	Polytomous C-R
HG000F1	Polytomous C-R
HH000A1	Polytomous C-R
HH000B1	Polytomous C-R
HH000C1	Polytomous C-R
HH000D1	Polytomous C-R
HH000F1	Polytomous C-R
HI000A1	Polytomous C-R
HI000B1	Polytomous C-R
HI000C1	Polytomous C-R
HI000D1	Polytomous C-R
HI000E1	Polytomous C-R
HI000F1	Polytomous C-R

Table E-6
NAEP 1997 Arts Assessment
Items Contributing to the
Theatre Responding Scale

Item	Item Type
HC00001	Multiple-choice
HC00002	Polytomous C-R
HC00004	Polytomous C-R
HC00006	Polytomous C-R
HC00008	Polytomous C-R
HC00009	Polytomous C-R
HD00001	Multiple-choice
HD00002	Multiple-choice
HD00004	Polytomous C-R
HD00009	Polytomous C-R
HE00001	Polytomous C-R
HE00002	Polytomous C-R
HE00003	Polytomous C-R
HE00005	Polytomous C-R
HE00006	Polytomous C-R
HE00007	Polytomous C-R
HF000B1	Polytomous C-R
HF00002	Polytomous C-R
HF00003	Multiple-choice
HF00004	Polytomous C-R
HF00005	Polytomous C-R
HF00007	Polytomous C-R
HF00008	Multiple-choice
HF00009	Polytomous C-R
HF00010	Multiple-choice
HF00011	Polytomous C-R
HF00B12	Polytomous C-R

Table E-7
NAEP 1997 Arts Assessment
Items Contributing to the
Visual Arts Creating Scale

Item	Item Type
VC000A6	Dichotomous C-R
VC000B6	Polytomous C-R
VE00009	Polytomous C-R
VF000B6	Polytomous C-R
VF000C6	Polytomous C-R
VF000B7	Dichotomous C-R
VF000C7	Polytomous C-R
VF000B8	Polytomous C-R
VG000A1	Polytomous C-R
VG000B1	Polytomous C-R
VG000C1	Polytomous C-R
VG000D1	Polytomous C-R
VG000E1	Polytomous C-R
VG000F1	Polytomous C-R
VG000G1	Polytomous C-R
VH00001	Polytomous C-R
VH000A2	Polytomous C-R
VH000B2	Polytomous C-R
VH000C2	Polytomous C-R
VH000A3	Polytomous C-R
VH000B3	Polytomous C-R
VH000C3	Polytomous C-R
VH000D3	Polytomous C-R
VI00001	Polytomous C-R
VI00002	Polytomous C-R
VI000A3	Polytomous C-R
VI000B3	Polytomous C-R
VI000C3	Polytomous C-R
VI000F3	Polytomous C-R
VI000G3	Polytomous C-R
VI000H3	Polytomous C-R
VI000I3	Polytomous C-R

Table E-8
NAEP 1997 Arts Assessment
Items Contributing to the
Visual Arts Responding Scale

Item	Item Type
VC00002	Polytomous C-R
VC00003	Multiple-choice
VCCL004	Cluster
VC00005	Polytomous C-R
VC00007	Polytomous C-R
VD00001	Dichotomous C-R
VD00002	Multiple-choice
VD00003	Multiple-choice
VD000A5	Polytomous C-R
VD000B5	Polytomous C-R
VD000A6	Polytomous C-R
VD000B6	Dichotomous C-R
VD00007	Polytomous C-R
VD00008	Polytomous C-R
VD00009	Polytomous C-R
VD00010	Multiple-choice
VD00011	Multiple-choice
VE00001	Dichotomous C-R
VE00002	Polytomous C-R
VE00003	Polytomous C-R
VE00004	Polytomous C-R
VE00005	Multiple-choice
VE00006	Multiple-choice
VE00007	Multiple-choice
VE00008	Polytomous C-R
VE00010	Polytomous C-R
VF00001	Polytomous C-R
VF00002	Multiple-choice
VF00003	Multiple-choice
VF00004	Polytomous C-R
VF000A6	Polytomous C-R
VF000A7	Polytomous C-R

Appendix F

NAEP 1997 ARTS ASSESSMENT IRT PARAMETERS

This appendix contains tables of IRT (item response theory) parameters for the NAEP 1997 Arts Assessment items that were scaled in each discipline for which IRT scales were created—Music, Theatre, and Visual Arts.

For each of the binary scored items used in scaling (i.e., multiple-choice items and short constructed-response items), the tables provide estimates of the IRT parameters and their associated standard errors (s.e.) of the estimates. For each of the polytomously scored items (i.e., the extended constructed-response items), the tables also show the estimates of the d_{jv} parameters which define category transition points (see equation 11.3 in Allen, Carlson, & Zelenak, 1998) and their associated standard errors.

Note that item parameters shown in this appendix are in the metrics used for the original calibration of the scales. The transformations needed to represent these parameters in terms of the metric of the final reporting scales are given in Section 3 (Music), Section 4 (Theatre), and Section 5 (Visual Arts).

Table F-1
1997 NAEP Arts Assessment IRT Parameters for Music Scale 1 Sample

NAEP ID	a_j(s.e.)	b_j(s.e.)	c_j(s.e.)	d_{j1}(s.e.)	d_{j2}(s.e.)	d_{j3}(s.e.)	d_{j4}(s.e.)
UCCL002	0.413 (-0.043)	1.926 (-0.198)	0.000 (0.000)				
UCCL003	0.330 (-0.015)	-0.652 (-0.065)	0.000 (0.000)	-3.822 (-0.254)	3.822 (-0.247)		
UC00004	0.836 (-0.055)	-0.453 (-0.051)	0.000 (0.000)				
UC00005	1.290 (-0.152)	0.160 (-0.086)	0.346 (-0.034)				
UC00006	1.032 (-0.062)	-0.055 (-0.041)	0.000 (0.000)				
UC00007	1.872 (-0.196)	0.581 (-0.047)	0.190 (-0.019)				
UCCL008	0.427 (-0.014)	-1.028 (-0.077)	0.000 (0.000)	-0.181 (-0.319)	3.841 (-0.234)	-3.661 (-0.117)	
UC00009	1.066 (-0.349)	3.259 (-0.483)	0.163 (-0.012)				
UC00010	0.690 (-0.040)	1.124 (-0.052)	0.000 (0.000)	1.009 (-0.057)	-1.009 (-0.100)		
UC00011	2.278 (-0.212)	1.077 (-0.049)	0.295 (-0.016)				
UC00012	1.249 (-0.076)	0.789 (-0.044)	0.000 (0.000)				
UC00013	1.101 (-0.173)	1.209 (-0.091)	0.266 (-0.024)				
UCCL014	0.547 (-0.023)	-0.252 (-0.061)	0.000 (0.000)	1.939 (-0.098)	-1.939 (-0.085)		
UCCL015	0.307 (-0.026)	4.056 (-0.192)	0.000 (0.000)	2.551 (-0.133)	-1.210 (-0.376)	1.407 (-0.657)	-2.748 (-2.226)
UD00004	0.474 (-0.108)	1.689 (-0.269)	0.301 (-0.047)				
UDCL005	0.535 (-0.066)	3.630 (-0.312)	0.000 (0.000)	0.505 (-0.127)	-0.505 (-0.468)		
UDCL006	0.222 (-0.018)	3.581 (-0.319)	0.000 (0.000)	-2.971 (-0.303)	2.971 (-0.413)		
UD00008	0.811 (-0.090)	-0.703 (-0.159)	0.298 (-0.056)				
UD00010	1.140 (-0.130)	0.251 (-0.091)	0.282 (-0.035)				
UD00011	1.077 (-0.155)	0.489 (-0.115)	0.401 (-0.036)				
UE00004	1.085 (-0.057)	0.144 (-0.029)	0.000 (0.000)	0.271 (-0.047)	-0.271 (-0.050)		
UE00005	1.878 (-0.224)	0.383 (-0.060)	0.314 (-0.025)				
UECL006	1.333 (-0.066)	0.780 (-0.029)	0.000 (0.000)	-0.127 (-0.050)	0.127 (-0.057)		
UE00007	2.299 (-0.312)	0.680 (-0.056)	0.375 (-0.021)				
UFCL002	0.504 (-0.071)	3.220 (-0.397)	0.000 (0.000)				
UFCL0A6	0.476 (-0.036)	3.108 (-0.135)	0.000 (0.000)	0.826 (-0.112)	1.354 (-0.185)	-0.856 (-0.486)	-1.323 (-1.478)
UFCL0B6	0.304 (-0.022)	2.829 (-0.105)	0.000 (0.000)	1.660 (-0.144)	1.487 (-0.195)	0.980 (-0.264)	-0.835 (-0.479)
					-1.999(-1.158)	-0.751 (0.000)	0.285 (-0.018)
UFCL0C6	1.513 (-0.169)	0.112 (-0.255)	-1.027 (-0.995)	-2.400 (-2.257)			
UFCL0D6	0.467 (-0.043)	2.753 (-0.173)	0.000 (0.000)	0.914 (-0.100)	-1.113 (-0.308)	0.198 (-0.571)	
UF00007	0.324 (-0.053)	-1.087 (-0.406)	0.312 (-0.066)				
UF00008	0.471 (-0.076)	0.090 (-0.271)	0.310 (-0.060)				

Table F-2
1997 NAEP Arts Assessment IRT Parameters for Music Scale 2 Sample

NAEP ID	a_j(s.e.)	b_j(s.e.)	c_j(s.e.)	d_{j1}(s.e.)	d_{j2}(s.e.)	d_{j3}(s.e.)
UCCL001	0.610 (0.055)	2.679 (0.162)	0.000 (0.000)	0.464 (0.090)	-0.464 (0.247)	
UD00001	0.537 (0.076)	0.567 (0.168)	0.194 (0.044)			
UD00002	0.472 (0.074)	0.656 (0.207)	0.211 (0.049)			
UD00003	0.549 (0.074)	0.574 (0.157)	0.182 (0.042)			
UD000A5	0.683 (0.062)	-1.905 (0.140)	0.000 (0.000)			
UD000B5	0.627 (0.049)	-0.764 (0.078)	0.000 (0.000)			
UD000C5	0.743 (0.052)	0.291 (0.056)	0.000 (0.000)			
UD000A6	0.487 (0.033)	-0.014 (0.053)	0.000 (0.000)	0.504 (0.093)	-0.504 (0.092)	
UD000B6	0.367 (0.021)	-0.180 (0.061)	0.000 (0.000)	-1.034 (0.141)	1.034 (0.137)	
UD000C6	0.630 (0.052)	1.117 (0.092)	0.000 (0.000)			
UD00007	0.702 (0.054)	0.906 (0.075)	0.000 (0.000)			
UDCL009	0.577 (0.046)	0.254 (0.069)	0.000 (0.000)			
UECL001	0.749 (0.099)	-2.942 (0.287)	0.000 (0.000)			
UE00002	0.303 (0.052)	-0.410 (0.383)	0.251 (0.063)			
UE00003	0.348 (0.052)	-1.721 (0.365)	0.235 (-0.113)			
UECL009	0.569 (0.027)	-0.984 (0.045)	0.000 (0.000)	-0.438 (0.141)	-0.073 (0.138)	0.511 (0.097)
UE00010	0.558 (0.035)	-1.020 (0.061)	0.000 (0.000)	1.087 (0.115)	-1.087 (0.073)	
UECL011	0.888 (0.062)	-1.860 (0.068)	0.000 (0.000)	0.794 (0.143)	-0.794 (0.055)	
UECL012	0.626 (0.043)	-1.574 (0.073)	0.000 (0.000)	0.906 (0.136)	-0.906 (0.067)	
UE00014	0.520 (0.093)	1.283 (0.190)	0.199 (0.042)			
UE00015	0.488 (0.021)	0.421 (0.075)	0.000 (0.000)	2.337 (0.102)	-2.337 (0.127)	
UE00016	0.523 (0.033)	-0.950 (0.068)	0.000 (0.000)	-0.274 (0.121)	0.274 (0.095)	
UF00003	0.560 (0.036)	-0.541 (0.052)	0.000 (0.000)	0.729 (0.092)	-0.729 (0.074)	
UF00004	0.444 (0.088)	0.993 (0.256)	0.263 (0.054)			
UF00005	0.420 (0.039)	4.155 (0.211)	0.000 (0.000)	1.411 (0.126)	1.611 (0.245)	-3.021 (1.702)

TableF-3
1997 NAEP Arts Assessment IRT Parameters for the Theatre Sample

NAEP ID	a_j(s.e.)	b_j(s.e.)	c_j(s.e.)	d_{j1}(s.e.)	d_{j2}(s.e.)	d_{j3}(s.e.)
HC00001	0.431 (0.084)	-3.849 (0.690)	0.216 (0.059)			
HC00002	0.732 (0.068)	-1.010 (0.103)	0.000 (0.000)			
HC00004	0.800 (0.060)	0.709 (0.052)	0.000 (0.000)	0.723 (0.069)	-0.723 (0.092)	
HC00006	0.452 (0.026)	1.636 (0.108)	0.000 (0.000)	2.939 (0.118)	-2.939 (0.297)	
HC00008	0.850 (0.055)	0.512 (0.051)	0.000 (0.000)	1.057 (0.069)	-1.057 (0.086)	
HC00009	0.447 (0.032)	-0.165 (0.068)	0.000 (0.000)	-0.536 (0.145)	0.536 (0.137)	
HD00001	0.814 (0.095)	-0.332 (0.125)	0.170 (0.043)			
HD00002	1.183 (0.133)	-0.071 (0.086)	0.167 (0.036)			
HD00004	0.393 (0.031)	-2.014 (0.166)	0.000 (0.000)	-1.026 (0.262)	1.026 (0.186)	
HD00009	0.842 (0.123)	2.125 (0.218)	0.000 (0.000)			
HE00001	0.498 (0.057)	-1.108 (0.150)	0.000 (0.000)			
HE00002	0.589 (0.060)	-0.327 (0.091)	0.000 (0.000)			
HE00003	0.663 (0.059)	1.156 (0.076)	0.000 (0.000)	0.575 (0.081)	-0.575 (0.128)	
HE00005	0.769 (0.065)	-1.281 (0.082)	0.000 (0.000)	0.555 (0.141)	-0.555 (0.077)	
HE00006	0.875 (0.078)	-0.363 (0.070)	0.000 (0.000)			
HE00007	0.539 (0.041)	-0.163 (0.074)	0.000 (0.000)	1.258 (0.127)	-1.258 (0.108)	
HF000B1	0.566 (0.041)	1.474 (0.074)	0.000 (0.000)	1.633 (0.093)	-0.580 (0.146)	-1.054 (0.316)
HF00002	0.448 (0.034)	0.300 (0.062)	0.000 (0.000)	1.333 (0.136)	-0.682 (0.141)	-0.651 (0.175)
HF00003	0.979 (0.133)	-0.216 (0.125)	0.244 (0.046)			
HF00004	1.070 (0.133)	1.613 (0.132)	0.000 (0.000)			
HF00005	0.388 (0.024)	2.984 (0.116)	0.000 (0.000)	3.996 (0.129)	-2.177 (0.370)	-1.820 (1.416)
HF00007	0.683 (0.035)	-0.291 (0.082)	0.000 (0.000)	2.261 (0.130)	-2.261 (0.108)	
HF00008	0.305 (0.058)	-3.243 (0.643)	0.223 (0.060)			
HF00009	0.612 (0.035)	-0.779 (0.081)	0.000 (0.000)	2.179 (0.159)	-2.179 (0.097)	
HF00010	0.881 (0.108)	-1.424 (0.173)	0.215 (0.053)			
HF00011	0.668 (0.038)	0.371 (0.068)	0.000 (0.000)	1.616 (0.092)	-1.616 (0.115)	
HF00B12	0.873 (0.061)	0.163 (0.047)	0.000 (0.000)	0.803 (0.071)	-0.803 (0.074)	

Table F-4
1997 NAEP Arts Assessment IRT Parameters for the Visual Arts Sample

NAEP ID	a_j (s.e.)	b_j (s.e.)	c_j (s.e.)	d_{j1} (s.e.)	d_{j2} (s.e.)	d_{j3} (s.e.)	d_{j4} (s.e.)
VC00002	0.627 (0.040)	-0.085 (0.044)	0.000 (0.000)	0.491 (0.077)	-0.491 (0.073)		
VC00003	0.291 (0.084)	3.151 (0.701)	0.215 (0.044)				
VCCL004	0.425 (0.021)	-0.011 (0.036)	0.000 (0.000)	-0.464 (0.150)	0.668 (0.159)	-0.310 (0.147)	0.107(0.136)
VC00005	0.656 (0.028)	0.514 (0.049)	0.000 (0.000)	2.626 (0.098)	-0.576 (0.071)	-2.050 (0.157)	
VC00007	0.509 (0.031)	0.937 (0.052)	0.000 (0.000)	0.012 (0.111)	0.844 (0.124)	-0.856 (0.150)	
VD00001	0.431 (0.045)	-1.046 (0.132)	0.000 (0.000)				
VD00002	0.318 (0.059)	0.598 (0.319)	0.211 (0.055)				
VD00003	0.260 (0.088)	5.381 (1.566)	0.179 (0.030)				
VD000A5	0.619 (0.035)	-0.124 (0.050)	0.000 (0.000)	1.138 (0.083)	-1.138 (0.076)		
VD000B5	0.563 (0.038)	-0.388 (0.052)	0.000 (0.000)	0.571 (0.092)	-0.571 (0.078)		
VD000A6	0.589 (0.041)	0.966 (0.062)	0.000 (0.000)	0.435 (0.075)	-0.435 (0.104)		
VD000B6	0.676 (0.053)	0.198 (0.061)	0.000 (0.000)				
VD00007	0.524 (0.027)	0.699 (0.064)	0.000 (0.000)	1.675 (0.084)	-1.675 (0.119)		
VD00008	0.503 (0.022)	1.078 (0.108)	0.000 (0.000)	3.527 (0.114)	-3.527 (0.242)		
VD00009	0.339 (0.022)	-1.488 (0.075)	0.000 (0.000)	3.021 (0.331)	-0.864 (0.140)	-2.157 (0.138)	
VD00010	0.993 (0.238)	1.836 (0.164)	0.190 (0.022)				
VD00011	0.503 (0.096)	1.026 (0.212)	0.232 (0.048)				
VE00001	0.317 (0.042)	-1.629 (0.241)	0.000 (0.000)				
VE00002	0.339 (0.045)	3.046 (0.342)	0.000 (0.000)	0.266 (0.154)	-0.266 (0.333)		
VE00003	0.632 (0.031)	1.701 (0.056)	0.000 (0.000)	3.334 (0.092)	0.345 (0.081)	-3.679 (0.689)	
VE00004	0.453 (0.048)	-1.244 (0.149)	0.000 (0.000)				
VE00005	0.727 (0.126)	1.303 (0.136)	0.166 (0.031)				
VE00006	0.598 (0.101)	0.772 (0.167)	0.225 (0.045)				
VE00007	0.977 (0.159)	1.302 (0.101)	0.150 (0.024)				
VE00008	0.657 (0.044)	0.721 (0.053)	0.000 (0.000)	0.850 (0.069)	-0.850 (0.094)		
VE00010	0.283 (0.047)	-4.927 (0.702)	0.000 (0.000)	0.702 (0.767)	-0.702 (0.236)		
VF00001	0.251 (0.020)	-0.751 (0.121)	0.000 (0.000)	1.907 (0.218)	-1.907 (0.182)		
VF00002	0.472 (0.079)	0.052 (0.239)	0.243 (0.056)				
VF00003	0.711 (0.094)	0.099 (0.136)	0.201 (0.044)				
VF00004	0.470 (0.030)	-0.732 (0.073)	0.000 (0.000)	1.504 (0.134)	-1.504 (0.098)		
VF000A6	0.303 (0.029)	1.378 (0.148)	0.000 (0.000)	-0.355 (0.162)	0.355 (0.212)		
VF000A7	0.291 (0.030)	0.328 (0.095)	0.000 (0.000)	0.421 (0.163)	-0.421 (0.177)		

Appendix G

NAEP 1997 ARTS ASSESSMENT SUMMARY TABLES OF VARIABLES USED TO DEFINE GROUPS OF STUDENTS

This appendix contains information about the variables used to define groups of students in the marginal maximum likelihood estimation of group score distributions for the NAEP 1997 Arts Assessment. The initial step in the construction of these variables involves forming primary student-based vectors of response data from answers to student, teacher, and school questionnaires, demographic and background data such as supplied by Westat, and other student information known prior to scaling. The initial vectors defining group membership concatenate this student background information into a series of identifying “contrasts” comprising:

1. Categorical variables derived by expanding the response options of a questionnaire variable into a binary series of one-degree-of-freedom “dummy” variables or contrasts (these form the majority of each student vector defining group membership);
2. Questionnaire or demographic variables that possess ordinal response options, such as number of hours spent watching television, which are included as linear and/or quadratic multidegree-of-freedom contrasts;
3. Continuous variables, such as student logit scores based on percent correct values, included as contrasts in their original form or a transformation of their original form; and
4. Interactions of two or more categorical variables forming a set of orthogonal one-degree-of-freedom dummy variables or contrasts.

This appendix contains lists of the variables used to define group membership. Table G-1 defines the information provided for each variable.

The linear model employed for the estimation of group score distributions did not directly use the variable specifications listed in this appendix. To eliminate inherent instabilities in estimation encountered when using a large number of correlated variables, a principal component transformation of the correlation matrix obtained from the variable contrasts derived according to these primary specifications was performed. The principal components scores based on this transformation were used as the predictor variables in estimating the linear model used to estimate group score distributions.

Table G-1
*NAEP 1997 Arts Assessment Description of Specifications
 Provided for Each Variable Defining Group Membership*

Title	Description
Grouping Variable ID	The eight-character NAEP database identification for the variable
NAEP ID	A unique eight-character ID assigned to identify each variable corresponding to a particular background or subject-area question within the entire pool of grouping variables. The first four characters identify the origin of the variable: BACK (background questionnaire), SUBJ (student questionnaire about the specific subject being assessed), SCHL (school questionnaire), TCHR (background part of teacher questionnaire), and TSUB (subject classroom part of teacher questionnaire). The second four digits represent the sequential position within each group of variables.
Description	A short description of the variable

Table G-2
Summary Table of the 1997 Music Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
BKSER	BACK0001	GRAND MEAN
DSEX	BACK0002	DERIVED SEX
DRACE	BACK0003	DERIVED RACE
NAEPRGN	BACK0004	NAEP REGION (1:NE;2:SE;3:CENT;4:WEST)
SCHTYP7	BACK0005	SCHOOL TYPE
TOL8	BACK0006	MSA/NON-MSA
TOL5	BACK0007	TYPE OF LOCALE (5 CATEGORIES)
DOC	BACK0008	DESCRIPTION OF COMMUNITY
PARED	BACK0009	PARENTS' HIGHEST LEVEL OF EDUCATION
HOMEEN	BACK0010	NUMBER OF ITEMS IN THE HOME (NEWSPAPER, > 25 BOOKS, ENCYCLOPEDIA, MAGAZINES)(DERIVED)
IEP	BACK0011	INDIVIDUALIZED EDUCATION PLAN
LEP	BACK0012	LIMITED ENGLISH PROFICIENCY
MUSELIG	BACK0013	ELIGIBLE FOR MUSIC
MUSADV	BACK0014	MUSIC EDVANCED
B000002	BACK0015	ARE YOU HISPANIC
B000003	BACK0016	HOW LONG HAVE YOU LIVED IN UNITED STATES
B000004	BACK0017	HOW OFTEN ENGLISH SPOKEN AT HOME
B000005	BACK0018	HOW FAR IN SCHOOL DID YOUR MOTHER GO
B000006	BACK0019	HOW FAR IN SCHOOL DID YOUR FATHER GO
B000007	BACK0020	DOES YOUR FAMILY GET A NEWSPAPER REGULARLY
B000008	BACK0021	IS THERE AN ENCYCLOPEDIA IN YOUR HOME
B000009	BACK0022	HOW MANY BOOKS IN YOUR HOME?
B000010	BACK0023	DOES YOUR FAMILY GET MAGAZINES REGULARLY
B000011	BACK0024	HOW MUCH TELEVISION DO YOU USUALLY WATCH EACH DAY
B000012	BACK0025	HOW MUCH TIME SPENT OF HOMEWORK EACH DAY
B000013	BACK0026	HOW OFTEN DO YOU READ FOR FUN ON OWN TIME
B000014	BACK0027	HOW MANY DAYS OF SCHOOL DID YOU MISS LAST MONTH
B000015	BACK0028	DID YOU GO TO PRESCHOOL BEFORE KINDERGARTEN
B000016	BACK0029	HOW MANY TIMES IN 2 YRS CHANGED SCHOOLS - MOVED
B000017	BACK0030	WITHIN PAST 2 YRS DID YOU CHANGE SCHOOLS - NO MOVE
B000018	BACK0031	HOW OFTEN DO YOU DISCUSS STUDIES IN SCHOOL AT HOME

(continued)

Table G-2 (continued)
Summary Table of the 1997 Music Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
B000019	BACK0032	HOW OFTEN DO YOU USE A COMPUTER AT HOME FOR SCHOOL
B000020	BACK0033	HOME MANY PAGES A DAY READ IN SCHOOL & HOMEWORK
B000021	BACK0034	WHICH STATEMENT BEST DESCRIBES GRADES SINCE GR 6
B000022	BACK0035	HOW MUCH EDUCATION DO YOU EXPECT TO RECEIVE
BM00001	SUBJ0001	I LIKE TO LISTEN TO MUSIC
BM00002	SUBJ0002	I LIKE TO PLAY MUSIC ALOND OR WITH OTHERS
BM00003	SUBJ0003	I THINK I HAVE TALENT FOR MUSIC
BM00004	SUBJ0004	PEOPLE TELL ME I AM A GOOD MUSICIAN
BM00005	SUBJ0005	I LIKE TO PLAY MUSIC FOR OTHER PEOPLE
BM00006	SUBJ0006	I WOULD LIKE TO BE A MUSICIAN WHEN I GROW UP
BM00007	SUBJ0007	PLAY MUSIC FOR YOU TO LISTEN TO
BM00008	SUBJ0008	ASK YOU TO SING
BM00009	SUBJ0009	ASK YOU TO PLAY INSTRUMENTS
BM00010	SUBJ0010	ASK YOU TO WRITE DOWN MUSIC
BM00011	SUBJ0011	ASK YOU TO WORK ON GROUP ASSIGNMENTS
BM00012	SUBJ0012	ASK YOU TO MAKE UP YOUR OWN MUSIC
BM00013	SUBJ0013	PLAY IN A BAND
BM00014	SUBJ0014	PLAY IN AN ORCHESTRA
BM00015	SUBJ0015	SING IN A CHORUS OR CHOIR
BM00016	SUBJ0016	TAKE PRIVATE SINGING LESSONS
BM00017	SUBJ0017	TAKE PRIVATE LESSONS ON AN INSTRUMENT
BM00018	SUBJ0018	NONE OF THE ABOVE
BM00019	SUBJ0019	DO YOU HAVE YOUR OWN MUSICAL INSTRUMENT
BM00020	SUBJ0020	HOW MANY TIMES DID YOU GO WITH CLASS TO CONCERT
BM00021	SUBJ0021	PLAY WITH A GROUP, BAND, OR ORCHESTRA
BM00022	SUBJ0022	SING IN A GROUP, CHORUS, OR CHOIR
BM00023	SUBJ0023	PLAY A MUSICAL INSTRUMENT ON YOUR OWN
BM00024	SUBJ0024	TAKE PRIVATE LESSONS ON AN INSTRUMENT OR SINGING
BM00025	SUBJ0025	LISTEN TO A MUSICAL TAPE, CD, OR RECORD
BM00026	SUBJ0026	ENTER A MUSIC COMPETITION
BM00027	SUBJ0027	GO TO A SUMMER MUSIC PROGRAM
BM00028	SUBJ0028	READ A BOOK ABOUT MUSIC
BM00029	SUBJ0029	WATCH A VIDEO, OR TV PROGRAM ABOUT MUSIC
BM00030	SUBJ0030	TALK WITH YOUR FAMILY OR FRIENDS ABOUT MUSIC
BM00031	SUBJ0031	NONE OF THE ABOVE
BM00032	SUBJ0032	AT SCHOOL
BM00033	SUBJ0033	ON THE STREET OR OUTDOORS
BM00034	SUBJ0034	IN A THEATRE
BM00035	SUBJ0035	ON TELEVISION
BM00036	SUBJ0036	OTHER: SPECIFY
BM00037	SUBJ0037	NONE OF THE ABOVE
SQ00004	SCHL0001	VISUAL ARTS
SQ00005	SCHL0002	DANCE
SQ00006	SCHL0003	MUSIC
SQ00007	SCHL0004	THEATRE

(continued)

Table G-2 (continued)
Summary Table of the 1997 Music Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
SQ00008	SCHL0005	VISUAL ARTS
SQ00009	SCHL0006	COMPUTERS AVAILABLE IN CLASSROOM
SQ00010	SCHL0007	COMPUTERS IN LABORATORIES
SQ00011	SCHL0008	BRING COMPUTERS TO CLASSROOM
SQ00012	SCHL0009	SPECIAL FOCUS DANCE
SQ00013	SCHL0010	SPECIAL FOCUS MUSIC
SQ00014	SCHL0011	SPECIAL FOCUS THEATRE
SQ00015	SCHL0012	SPECIAL FOCUS VISUAL ARTS
SQ00016	SCHL0013	SPECIAL FOCUS OTHER
SQ00017	SCHL0014	NOT A SPECIAL FOCUS SCHOOL
SQ00018	SCHL0015	DANCE CURRICULUM
SQ00019	SCHL0016	MUSIC CURRICULUM
SQ00020	SCHL0017	THEATRE CURRICULUM
SQ00021	SCHL0018	CURRICULUM VISUAL ARTS
SQ00022	SCHL0019	CURRICULUM NONE OF THE ABOVE
SQ00023	SCHL0020	DANCE FIELD TRIPS
SQ00024	SCHL0021	MUSIC FIELD TRIPS
SQ00025	SCHL0022	THEATRE FIELD TRIPS
SQ00026	SCHL0023	VISUAL ARTS FIELD TRIPS
SQ00027	SCHL0024	OTHER (SPECIFY) FIELD TRIPS
SQ00028	SCHL0025	NONE OF THE ABOVE FIELD TRIPS
SQ00029	SCHL0026	EXTRACURRICULAR SPONSORED DANCE
SQ00030	SCHL0027	EXTRACURRICULAR SPONSORED MUSIC
SQ00031	SCHL0028	EXTRACURRICULAR SPONSORED THEATRE
SQ00032	SCHL0029	EXTRACURRICULAR SPONSORED VISUAL ARTS
SQ00033	SCHL0030	EXTRACURRICULAR SPONSORED NONE OF THE ABOVE
SQ00034	SCHL0031	SUMMER PROGRAM DANCE
SQ00035	SCHL0032	SUMMER PROGRAM MUSIC
SQ00036	SCHL0033	SUMMER PROGRAM THEATRE
SQ00037	SCHL0034	SUMMER PROGRAM VISUAL ARTS
SQ00038	SCHL0035	SUMMER PROGRAMS NONE OF THE ABOVE
SQ00039	SCHL0036	POSITION OF STAFF - DANCE FULL-TIME SPECIALIST
SQ00040	SCHL0037	POSITION OF STAFF - DANCE PART-TIME SPECIALIST
SQ00041	SCHL0038	POSITION OF STAFF- DANCE PHYSICAL ED TEACHER
SQ00042	SCHL0039	POSITION OF STAFF- DANCE ELEMENTARY TEACHER
SQ00043	SCHL0040	POSITION OF STAFF-OTHER FACULTY MEMBER
SQ00044	SCHL0041	POSITION OF STAFF- DANCE ARTIST IN RESIDENCE
SQ00045	SCHL0042	POSITION OF STAFF- DANCE VOLUNTEER
SQ00046	SCHL0043	POSITION OF STAFF-DANCE OR MOVEMENT NOT TAUGHT
SQ00047	SCHL0044	POSITION OF STAFF-MUSIC FULLTIME SPECIALIST
SQ00048	SCHL0045	POSITION OF STAFF-MUSIC PART TIME SPECIALIST
SQ00049	SCHL0046	POSITION OF STAFF-MUSIC ELEMENTARY TEACHER
SQ00050	SCHL0047	POSITION OF STAFF-MUSIC OTHER FACULTY MEMBER
SQ00051	SCHL0048	POSITION OF STAFF-MUSIC ARTIST IN RESIDENCE
SQ00052	SCHL0049	POSITIONOF STAFF-MUSIC VOLUNTEER

(continued)

Table G-2 (continued)
Summary Table of the 1997 Music Variables Defining Group Membership

Group Variable ID	NAEP ID	Description
SQ00053	SCHL0050	POSITION OF STAFF-MUSIC NOT TAUGHT
SQ00054	SCHL0051	POSITION OF STAFF -THEATRE FULL TIME SPECIALIST
SQ00055	SCHL0052	POSITION OF STAFF-THEATRE PART TIME SPECIALIST
SQ00056	SCHL0053	POSITION OF STAFF-THEATRE ENGLISH LANG ARTS TEACHE
SQ00057	SCHL0054	POSITION OF STAFF-THEATRE ELEMENTARY TEACHER
SQ00058	SCHL0055	POSITION OF STAFF-THEATRE OTHER FACULTY MEMBER
SQ00059	SCHL0056	POSITION OF STAFF-THEATRE ARTIST IN RESIDENCE
SQ00060	SCHL0057	POSITION OF STAFF-THEATRE VOLUNTEER
SQ00061	SCHL0058	POSITION OF STAFF-THEATRE OR DRAMA NOT TAUGHT
SQ00062	SCHL0059	POSITION OF STAFF-VISUAL ARTS FULLTIME SPECIALIST
SQ00063	SCHL0060	POSITION OF STAFF-VISUAL ARTS PART TIME SPECIALIST
SQ00064	SCHL0061	POSITION OF STAFF-VISUAL ARTS ELEMENTARY TEACHER
SQ00065	SCHL0062	POSITION OF STAFF-VISUAL ARTS OTHER FACULTY MEMBER
SQ00066	SCHL0063	POSITIONOF STAFF-VISUAL ARTS ARTIST IN RESIDENCE
SQ00067	SCHL0064	POSITION OF STAFF-VISUAL ARTS VOLUNTEER
SQ00068	SCHL0065	POSITION OF STAFF-VISUAL ARTS NOT TAUGHT
SQ00069	SCHL0066	WHICH BEST DESCRIBES SPACE FOR TEACHING DANCE
SQ00070	SCHL0067	WHICH BEST DESCRIBES SPACE FOR PERFORMING MUSIC
SQ00071	SCHL0068	WHICH BEST DESCRIBES SPACE FOR TEACHING THEATRE
SQ00072	SCHL0069	WHICH BEST DESCRIBES SPACE FOR TEACHING VISUAL ART
SQ00073	SCHL0070	DID SCHOOL BRING VISITING ARTISTS - DANCE
SQ00074	SCHL0071	DID SCHOOL BRING VISITING ARTISTS - MUSIC
SQ00075	SCHL0072	DID SCHOOL BRING VISITING ARTISTS - THEATRE
SQ00076	SCHL0073	DID SCHOOL BRING VISITING ARTISTS - VISUAL ARTS
SQ00077	SCHL0074	DID SCHOOL BRING VISTING ARTISTS - NONE
SQ00078	SCHL0075	DID SCHOOL SPONSOR VISITING ARTIST PROG -DANCE
SQ00079	SCHL0076	DID SCHOOL SPONSOR VISITING ARTIST PROG - FILM
SQ00080	SCHL0077	DID SCHOOL SPONSOR VISITING ARTIST PROG-FOLK ARTS
SQ00081	SCHL0078	DID SCHOOL SPONSOR VISITING ARTIST PROG-MUSIC
SQ00082	SCHL0079	DID SCHOOL SPONSOR VISITING ARTISTS PROG-THEATRE
SQ00083	SCHL0080	DID SCHOOL SPONSOR VISITING ARTIST PROG-VISUAL ART
SQ00084	SCHL0081	DID SCHOOL SPONSOR VISITING ARTIST PROG-NONE
SQ00085	SCHL0082	WHAT PERCENT 8TH GRADE INSTRUCTED - DANCE
SQ00086	SCHL0083	WHAT PERCENT 8TH GRADE INSTRUCTED - MUSIC
SQ00087	SCHL0084	WHAT PERCENT 8TH GRADE INSTRUCTED - THEATRE
SQ00088	SCHL0085	WHAT PERCENT 8TH GRADE INSTRUCTED - VISUAL ARTS
SQ00089	SCHL0086	WHAT IS THE CURRENT ENROLLMENT IN YOUR SCHOOL?
SQ00090	SCHL0087	WHAT IS THE CURRENT ENROLLMENT IN THE EIGHTH GRADE
SQ00091	SCHL0088	DOES SCHOOL PARTICIPATE-NATIONAL SCHOOL LUNCH PROG
SQ00092	SCHL0089	TOTAL # STUDENTS ELIGIBLE-NATL SCHOOL LUNCH PROG
SQ00093	SCHL0090	HOW MANY 8TH GRADE ELIGIBLE NATL SCHOOL LUNCH PROG
SQ00094	SCHL0091	DOES SCHOOL RECEIVE CHPATER1/TITLE 1 FUNDING
SQ00095	SCHL0092	TOTAL # STUDENTS ELIGIBLE CHAPTER 1/TITLE 1
SQ00096	SCHL0093	HOW MANY 8TH GRADE ELIGIBLE CHAPTER 1/TITLE 1
SQ00097	SCHL0094	WHAT PERCENT STUDENTS ELIGIBLE CHAPTER1/TITLE 1

(continued)

Table G-2 (continued)
Summary Table of the 1997 Music Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
SQ00A98	SCHL0095	TOTAL SCHOOL NATIONAL SCHOOL LUNCH PROGRAM
SQ00B98	SCHL0096	TOTAL EIGHTH GRADE NATIONAL SCHOOL LUNCH PROGRAM
SQ00A99	SCHL0097	TOTAL SCHOOL CHAPTER 1/TITLE 1
SQ00B99	SCHL0098	TOTAL EIGHTH GRADE CHAPTER1/TITLE 1
SQ0A100	SCHL0099	TOTAL SCHOOL REMEDIAL READING INSTRUCTION
SQ0B100	SCHL0100	TOTAL EIGHTH GRADE REMEDIAL READING INSTRUCTION
SQ0A101	SCHL0101	TOTAL SCHOOL REMEDIAL MATH INSTRUCTION
SQ0B101	SCHL0102	TOTAL EIGHTH GRADE REMEDIAL MATH INSTRUCTION
SQ0A102	SCHL0103	TOTAL SCHOOL BILINGUAL EDUCATION
SQ0B102	SCHL0104	TOTAL EIGHTH GRADE BILINGUAL EDUCATION
SQ0A103	SCHL0105	TOTAL SCHOOL ENGLISH AS SECOND LANGUAGE INSTRUCT
SQ0B103	SCHL0106	EIGHTH GRADE ENGLISH AS SECOND LANGUAGE INSTRUCT
SQ0A104	SCHL0107	TOTAL SCHOOL SPECIAL EDUCATION PARTICIPANTS
SQ0B104	SCHL0108	TOTAL EIGHTH GRADE SPECIAL EDUCATION PARTICIPANTS
SQ0A105	SCHL0109	TOTAL SCHOOL GIFTED AND TALENTED EDUCATION PARTICI
SQ0B105	SCHL0110	TOTAL EIGHTH GRADE GIFTED AND TALENTED EDUCATION
SQ00106	SCHL0111	WHAT PERCENT OF PARENTS PARTICIPATES IN PTO
SQ00107	SCHL0112	WHAT PERCENT OF PARENTS PARTICIPATES IN OPEN HOUSE
SQ00108	SCHL0113	WHAT PERCENT OF PARENTS PARTICIPATE IN CONFERENCES
SQ00109	SCHL0114	WHAT PERCENT OF PARENTS INVOLVED IN CURRICULUM DEC
SQ00110	SCHL0115	WHAT PERCENT OF PARENTS PARTICIPATES IN VOLUNTEER
SQ00111	SCHL0116	WHAT DEGREE IS STUDENT ABSENTEEISM A PROBLEM
SQ00112	SCHL0117	WHAT DEGREE IS STUDENT TARDINESS A PROBLEM
SQ00113	SCHL0118	WHAT DEGREE IS PHYSICAL CONFLICTS/CHILDREN A PROBL
SQ00114	SCHL0119	WHAT DEGREE IS TEACHER ABSENTEEISM A PROBLEM
SQ00115	SCHL0120	WHAT DEGREE IS RACIAL/CULTURAL CONFLICTS A PROBLEM
SQ00116	SCHL0121	WHAT DEGREE IS STUDENT HEALTH A PROBLEM
SQ00117	SCHL0122	WHAT DEGREE IS LACK OF PARENT INVOLVEMENT A PROBLE
SQ00118	SCHL0123	WHAT DEGREE IS STUDENT USE OF ALCOHOL A PROBLEM
SQ00119	SCHL0124	WHAT DEGREE IS STUDENT USE OF TOBACCO A PROBLEM
SQ00120	SCHL0125	WHAT DEGREE IS STUDENT USE OF DRUGS A PROBLEM
SQ00121	SCHL0126	WHAT DEGREE IS GANG ACTIVITY A PROBLEM
SQ00122	SCHL0127	WHAT DEGREE IS STUDENT MISBEHAVIOR IN CLASS A PROB
SQ00123	SCHL0128	WHAT DEGREE IS CHEATING A PROBLEM IN YOUR SCHOOL
SQ00124	SCHL0129	HOW CHARACTERIZE MORALE OF TEACHERS IN YOUR SCHOOL
SQ00125	SCHL0130	HOW CHARACTERIZE STUDENTS' ATTITUDES/ACADEM ACHIEV
SQ00126	SCHL0131	HOW CHARACTERIZE PARENTAL SUPPORT/STUDENT ACHIEVEM
SQ00127	SCHL0132	HOW CHARACTERIZE REGARD FOR SCHOOL PROPERTY
SQ00128	SCHL0133	PERCENTAGE OF STUDENTS ABSENT ON AN AVERAGE DAY
SQ00129	SCHL0134	WHAT PERCENT TEACHERS ABSENT AVERAGE DAY
SQ00130	SCHL0135	WHAT PERCENT STUDENT STILL ENROLLED AT END OF YEAR
SQ00131	SCHL0136	WHAT PERCENT THIS YEARS EIGHTH GRADERS HELD BACK
SQ00132	SCHL0137	WHAT PERCENT FULL-TIME STARTING TEACHERS LEFT

Table G-3
Summary Table of the 1997 Theatre Variables Defining Group Membership

Group Variable ID	NAEP ID	Description
BKSER	BACK0001	GRAND MEAN
DSEX	BACK0002	DERIVED SEX
DRACE	BACK0003	DERIVED RACE
NAEPRGN	BACK0004	NAEP REGION (1:NE;2:SE;3:CENT;4:WEST)
SCHTYP7	BACK0005	SCHOOL TYPE
TOL8	BACK0006	MSA/NON-MSA
TOL5	BACK0007	TYPE OF LOCALE (5 CATEGORIES)
DOC	BACK0008	DESCRIPTION OF COMMUNITY
PARED	BACK0009	PARENTS' HIGHEST LEVEL OF EDUCATION
HOMEEN	BACK0010	NUMBER OF ITEMS IN THE HOME (NEWSPAPER, > 25 BOOKS, ENCYCLOPEDIA, MAGAZINES)(DERIVED)
IEP	BACK0011	INDIVIDUALIZED EDUCATION PLAN
LEP	BACK0012	LIMITED ENGLISH PROFICIENCY
MUSELIG	BACK0013	ELIGIBLE FOR MUSIC
MUSADV	BACK0014	MUSIC EDVANCED
B000002	BACK0015	ARE YOU HISPANIC
B000003	BACK0016	HOW LONG HAVE YOU LIVED IN UNITED STATES
B000004	BACK0017	HOW OFTEN ENGLISH SPOKEN AT HOME
B000005	BACK0018	HOW FAR IN SCHOOL DID YOUR MOTHER GO
B000006	BACK0019	HOW FAR IN SCHOOL DID YOUR FATHER GO
B000007	BACK0020	DOES YOUR FAMILY GET A NEWSPAPER REGULARLY
B000008	BACK0021	IS THERE AN ENCYCLOPEDIA IN YOUR HOME
B000009	BACK0022	HOW MANY BOOKS IN YOUR HOME?
B000010	BACK0023	DOES YOUR FAMILY GET MAGAZINES REGULARLY
B000011	BACK0024	HOW MUCH TELEVISION DO YOU USUALLY WATCH EACH DAY
B000012	BACK0025	HOW MUCH TIME SPENT OF HOMEWORK EACH DAY
B000013	BACK0026	HOW OFTEN DO YOU READ FOR FUN ON OWN TIME
B000014	BACK0027	HOW MANY DAYS OF SCHOOL DID YOU MISS LAST MONTH
B000015	BACK0028	DID YOU GO TO PRESCHOOL BEFORE KINDERGARTEN
B000016	BACK0029	HOW MANY TIMES IN 2 YRS CHANGED SCHOOLS - MOVED
B000017	BACK0030	WITHIN PAST 2 YRS DID YOU CHANGE SCHOOLS - NO MOVE
B000018	BACK0031	HOW OFTEN DO YOU DISCUSS STUDIES IN SCHOOL AT HOME
B000019	BACK0032	HOW OFTEN DO YOU USE A COMPUTER AT HOME FOR SCHOOL
B000020	BACK0033	HOME MANY PAGES A DAY READ IN SCHOOL & HOMEWORK
B000021	BACK0034	WHICH STATEMENT BEST DESCRIBES GRADES SINCE GR 6
B000022	BACK0035	HOW MUCH EDUCATION DO YOU EXPECT TO RECEIVE
BT00001	SUBJ0001	I LIKE TO WATCH PLAYS AND MUSICALS
BT00002	SUBJ0002	I LIKE TO ACT IN OR HELP WITH PLAYS OR MUSICALS
BT00003	SUBJ0003	I LIKE TO WRITE PLAYS OR SCRIPTS
BT00004	SUBJ0004	I THINK I HAVE TALENT FOR DRAMA OR THEATRE
BT00005	SUBJ0005	PEOPLE TELL ME I AM A GOOD ACTOR OR DIRECTOR
BT00006	SUBJ0006	I LIKE OTHER PEOPLE TO SEE PLAYS I ACT OR WORK IN
BT00007	SUBJ0007	I'D LIKE TO BE IN THEATRE OR MOVIES WHEN I GROW UP
BT00008	SUBJ0008	LIVE PRODUCTIONS YOU HAVE ACTED/HELPED IN
BT00009	SUBJ0009	IN SCHOOL: INVOLVED IN ACTING

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Membership

Grouping Variable ID	NAEP ID	Description
BT00010	SUBJ0010	IN SCHOOL: INVOLVED IN TECHNICAL WORK
BT00011	SUBJ0011	IN SCHOOL: INVOLVED IN DIRECTING
BT00012	SUBJ0012	IN SCHOOL: INVOLVED IN WRITING A PLAY
BT00013	SUBJ0013	IN SCHOOL: I HAVE NOT BEEN INVOLVED
BT00014	SUBJ0014	OUT OF SCHOOL: INVOLVED IN ACTING
BT00015	SUBJ0015	OUT OF SCHOOL: INVOLVED IN TECHNICAL WORK
BT00016	SUBJ0016	OUT OF SCHOOL: INVOLVED IN DIRECTING
BT00017	SUBJ0017	OUT OF SCHOOL INVOLVED IN WRITING PLAY OR SCRIPT
BT00018	SUBJ0018	OUT OF SCHOOL: I HAVE NOT BEEN INVOLVED
BT00019	SUBJ0019	IN SCHOOL: PLAY ROLES
BT00020	SUBJ0020	IN SCHOOL: READ A PLAY OR A SCENE SILENTLY
BT00021	SUBJ0021	IN SCHOOL: READ A PLAY OR A SCENE ALOUD
BT00022	SUBJ0022	IN SCHOOL: IMPROVISE SCENES
BT00023	SUBJ0023	IN SCHOOL: WRITE A PLAY OR A SCENE
BT00024	SUBJ0024	IN SCHOOL: DO TECHNICAL THEATRE
BT00025	SUBJ0025	IN SCHOOL: PERFORM FOR AN AUDIENCE
BT00026	SUBJ0026	IN SCHOOL: READ ABOUT THEATRE
BT00027	SUBJ0027	IN SCHOOL: WATCH PLAY/THEATRE PROG ON TV/VIDEO
BT00028	SUBJ0028	CLASSROOM INSTRUCTION: MAKING A FILM OR VIDEO
BT00029	SUBJ0029	CLASSROOM INSTRUCTION: ACTING
BT00030	SUBJ0030	CLASSROOM INSTRUCTION: DIRECTING
BT00031	SUBJ0031	CLASSROOM INSTRUCTION: PLAYWRITING
BT00032	SUBJ0032	CLASSROOM INSTRUCTION: TECHNICAL THEATRE
BT00033	SUBJ0033	YOU/TEACHER SAVE DRAMA WORK PORTFOLIO/ON VIDEO
BT00A34	SUBJ0034	THIS YEAR, USED VIDEO/FILM CAMERA - YES, IN SCHOOL
BT00B34	SUBJ0035	THIS YEAR, USED VIDEO/FILM CAMERA - YES, AT HOME
BT00C34	SUBJ0036	THIS YEAR, USED VIDEO/FILM CAMERA - NO
BT00035	SUBJ0037	IN SCHOOL, MADE/DIRECTED MOVIE/DOCUMENTARY/VIDEO
BT00036	SUBJ0038	LIVE THEATRE PERFORMANCES ATTENDED LAST YEAR
BT00037	SUBJ0039	NOT FOR SCHOOL: GO TO PLAY, MUSICAL, PUPPET SHOW
BT00038	SUBJ0040	NOT FOR SCHOOL: TAKE ACTING/DRAMA CLASSES
BT00039	SUBJ0041	NOT FOR SCHOOL: SUMMER THEATRE PROGRAM
BT00040	SUBJ0042	NOT FOR SCHOOL: ENTER ACTING COMPETITION
BT00041	SUBJ0043	NOT FOR SCHOOL: WRITE A PLAY OR SCRIPT
BT00042	SUBJ0044	NOT FOR SCHOOL: LOOK AT/READ A BOOK ABOUT THEATRE
BT00043	SUBJ0045	NOT IN SCHOOL: WATCH PLAY/THEATRE PROG ON TV/VIDEO
BT00044	SUBJ0046	NOT FOR SCHOOL: TALK TO FAMILY/FRIENDS RE: THEATRE
BT00045	SUBJ0047	NOT FOR SCHOOL: NONE OF THE ABOVE
SQ00004	SCHL0001	VISUAL ARTS
SQ00005	SCHL0002	DANCE
SQ00006	SCHL0003	MUSIC
SQ00007	SCHL0004	THEATRE
SQ00008	SCHL0005	VISUAL ARTS
SQ00009	SCHL0006	COMPUTERS AVAILABLE IN CLASSROOM
SQ00010	SCHL0007	COMPUTERS IN LABORATORIES

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Group Variable ID	NAEP ID	Description
SQ00011	SCHL0008	BRING COMPUTERS TO CLASSROOM
SQ00012	SCHL0009	SPECIAL FOCUS DANCE
SQ00013	SCHL0010	SPECIAL FOCUS MUSIC
SQ00014	SCHL0011	SPECIAL FOCUS THEATRE
SQ00015	SCHL0012	SPECIAL FOCUS VISUAL ARTS
SQ00016	SCHL0013	SPECIAL FOCUS OTHER
SQ00017	SCHL0014	NOT A SPECIAL FOCUS SCHOOL
SQ00018	SCHL0015	DANCE CURRICULUM
SQ00019	SCHL0016	MUSIC CURRICULUM
SQ00020	SCHL0017	THEATRE CURRICULUM
SQ00021	SCHL0018	CURRICULUM VISUAL ARTS
SQ00022	SCHL0019	CURRICULUM NONE OF THE ABOVE
SQ00023	SCHL0020	DANCE FIELD TRIPS
SQ00024	SCHL0021	MUSIC FIELD TRIPS
SQ00025	SCHL0022	THEATRE FIELD TRIPS
SQ00026	SCHL0023	VISUAL ARTS FIELD TRIPS
SQ00027	SCHL0024	OTHER (SPECIFY) FIELD TRIPS
SQ00028	SCHL0025	NONE OF THE ABOVE FIELD TRIPS
SQ00029	SCHL0026	EXTRACURRICULAR SPONSORED DANCE
SQ00030	SCHL0027	EXTRACURRICULAR SPONSORED MUSIC
SQ00031	SCHL0028	EXTRACURRICULAR SPONSORED THEATRE
SQ00032	SCHL0029	EXTRACURRICULAR SPONSORED VISUAL ARTS
SQ00033	SCHL0030	EXTRACURRICULAR SPONSORED NONE OF THE ABOVE
SQ00034	SCHL0031	SUMMER PROGRAM DANCE
SQ00035	SCHL0032	SUMMER PROGRAM MUSIC
SQ00036	SCHL0033	SUMMER PROGRAM THEATRE
SQ00037	SCHL0034	SUMMER PROGRAM VISUAL ARTS
SQ00038	SCHL0035	SUMMER PROGRAMS NONE OF THE ABOVE
SQ00039	SCHL0036	POSITION OF STAFF - DANCE FULL-TIME SPECIALIST
SQ00040	SCHL0037	POSITION OF STAFF - DANCE PART-TIME SPECIALIST
SQ00041	SCHL0038	POSITION OF STAFF- DANCE PHYSICAL ED TEACHER
SQ00042	SCHL0039	POSITION OF STAFF- DANCE ELEMENTARY TEACHER
SQ00043	SCHL0040	POSITION OF STAFF-OTHER FACULTY MEMBER
SQ00044	SCHL0041	POSITION OF STAFF- DANCE ARTIST IN RESIDENCE
SQ00045	SCHL0042	POSITION OF STAFF- DANCE VOLUNTEER
SQ00046	SCHL0043	POSITION OF STAFF-DANCE OR MOVEMENT NOT TAUGHT
SQ00047	SCHL0044	POSITION OF STAFF-MUSIC FULLTIME SPECIALIST
SQ00048	SCHL0045	POSITION OF STAFF-MUSIC PART TIME SPECIALIST
SQ00049	SCHL0046	POSITION OF STAFF-MUSIC ELEMENTARY TEACHER
SQ00050	SCHL0047	POSITION OF STAFF-MUSIC OTHER FACULTY MEMBER
SQ00051	SCHL0048	POSITION OF STAFF-MUSIC ARTIST IN RESIDENCE
SQ00052	SCHL0049	POSITIONOF STAFF-MUSIC VOLUNTEER
SQ00053	SCHL0050	POSITION OF STAFF-MUSIC NOT TAUGHT
SQ00054	SCHL0051	POSITION OF STAFF -THEATRE FULL TIME SPECIALIST
SQ00055	SCHL0052	POSITION OF STAFF-THEATRE PART TIME SPECIALIST

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
SQ00056	SCHL0053	POSITION OF STAFF-THEATRE ENGLISH LANG ARTS TEACHE
SQ00057	SCHL0054	POSITION OF STAFF-THEATRE ELEMENTARY TEACHER
SQ00058	SCHL0055	POSITION OF STAFF-THEATRE OTHER FACULTY MEMBER
SQ00059	SCHL0056	POSITION OF STAFF-THEATRE ARTIST IN RESIDENCE
SQ00060	SCHL0057	POSITION OF STAFF-THEATRE VOLUNTEER
SQ00061	SCHL0058	POSITION OF STAFF-THEATRE OR DRAMA NOT TAUGHT
SQ00062	SCHL0059	POSITION OF STAFF-VISUAL ARTS FULLTIME SPECIALIST
SQ00063	SCHL0060	POSITION OF STAFF-VISUAL ARTS PART TIME SPECIALIST
SQ00064	SCHL0061	POSITION OF STAFF-VISUAL ARTS ELEMENTARY TEACHER
SQ00065	SCHL0062	POSITION OF STAFF-VISUAL ARTS OTHER FACULTY MEMBER
SQ00066	SCHL0063	POSITIONOF STAFF-VISUAL ARTS ARTIST IN RESIDENCE
SQ00067	SCHL0064	POSITION OF STAFF-VISUAL ARTS VOLUNTEER
SQ00068	SCHL0065	POSITION OF STAFF-VISUAL ARTS NOT TAUGHT
SQ00069	SCHL0066	WHICH BEST DESCRIBES SPACE FOR TEACHING DANCE
SQ00070	SCHL0067	WHICH BEST DESCRIBES SPACE FOR PERFORMING MUSIC
SQ00071	SCHL0068	WHICH BEST DESCRIBES SPACE FOR TEACHING THEATRE
SQ00072	SCHL0069	WHICH BEST DESCRIBES SPACE FOR TEACHING VISUAL ART
SQ00073	SCHL0070	DID SCHOOL BRING VISITING ARTISTS - DANCE
SQ00074	SCHL0071	DID SCHOOL BRING VISITING ARTISTS - MUSIC
SQ00075	SCHL0072	DID SCHOOL BRING VISITING ARTISTS - THEATRE
SQ00076	SCHL0073	DID SCHOOL BRING VISITING ARTISTS - VISUAL ARTS
SQ00077	SCHL0074	DID SCHOOL BRING VISTING ARTISTS - NONE
SQ00078	SCHL0075	DID SCHOOL SPONSOR VISITING ARTIST PROG -DANCE
SQ00079	SCHL0076	DID SCHOOL SPONSOR VISITING ARTIST PROG - FILM
SQ00080	SCHL0077	DID SCHOOL SPONSOR VISITING ARTIST PROG-FOLK ARTS
SQ00081	SCHL0078	DID SCHOOL SPONSOR VISITING ARTIST PROG-MUSIC
SQ00082	SCHL0079	DID SCHOOL SPONSOR VISITING ARTISTS PROG-THEATRE
SQ00083	SCHL0080	DID SCHOOL SPONSOR VISITING ARTIST PROG-VISUAL ART
SQ00084	SCHL0081	DID SCHOOL SPONSOR VISITING ARTIST PROG-NONE
SQ00085	SCHL0082	WHAT PERCENT 8TH GRADE INSTRUCTED - DANCE
SQ00086	SCHL0083	WHAT PERCENT 8TH GRADE INSTRUCTED - MUSIC
SQ00087	SCHL0084	WHAT PERCENT 8TH GRADE INSTRUCTED - THEATRE
SQ00088	SCHL0085	WHAT PERCENT 8TH GRADE INSTRUCTED - VISUAL ARTS
SQ00089	SCHL0086	WHAT IS THE CURRENT ENROLLMENT IN YOUR SCHOOL?
SQ00090	SCHL0087	WHAT IS THE CURRENT ENROLLMENT IN THE EIGHTH GRADE
SQ00091	SCHL0088	DOES SCHOOL PARTICIPATE-NATIONAL SCHOOL LUNCH PROG
SQ00092	SCHL0089	TOTAL # STUDENTS ELIGIBLE-NATL SCHOOL LUNCH PROG
SQ00093	SCHL0090	HOW MANY 8TH GRADE ELIGIBLE NATL SCHOOL LUNCH PROG
SQ00094	SCHL0091	DOES SCHOOL RECEIVE CHPATER1/TITLE 1 FUNDING
SQ00095	SCHL0092	TOTAL # STUDENTS ELIGIBLE CHAPTER 1/TITLE 1
SQ00096	SCHL0093	HOW MANY 8TH GRADE ELIGIBLE CHAPTER 1/TITLE 1
SQ00097	SCHL0094	WHAT PERCENT STUDENTS ELIGIBLE CHAPTER1/TITLE 1
SQ00A98	SCHL0095	TOTAL SCHOOL NATIONAL SCHOOL LUNCH PROGRAM
SQ00B98	SCHL0096	TOTAL EIGHTH GRADE NATIONAL SCHOOL LUNCH PROGRAM
SQ00A99	SCHL0097	TOTAL SCHOOL CHAPTER 1/TITLE 1

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
SQ00B99	SCHL0098	TOTAL EIGHTH GRADE CHAPTER1/TITLE 1
SQ0A100	SCHL0099	TOTAL SCHOOL REMEDIAL READING INSTRUCTION
SQ0B100	SCHL0100	TOTAL EIGHTH GRADE REMEDIAL READING INSTRUCTION
SQ0A101	SCHL0101	TOTAL SCHOOL REMEDIAL MATH INSTRUCTION
SQ0B101	SCHL0102	TOTAL EIGHTH GRADE REMEDIAL MATH INSTRUCTION
SQ0A102	SCHL0103	TOTAL SCHOOL BILINGUAL EDUCATION
SQ0B102	SCHL0104	TOTAL EIGHTH GRADE BILINGUAL EDUCATION
SQ0A103	SCHL0105	TOTAL SCHOOL ENGLISH AS SECOND LANGUAGE INSTRUCT
SQ0B103	SCHL0106	EIGHTH GRADE ENGLISH AS SECOND LANGUAGE INSTRUCT
SQ0A104	SCHL0107	TOTAL SCHOOL SPECIAL EDUCATION PARTICIPANTS
SQ0B104	SCHL0108	TOTAL EIGHTH GRADE SPECIAL EDUCATION PARTICIPANTS
SQ0A105	SCHL0109	TOTAL SCHOOL GIFTED AND TALENTED EDUCATION PARTICI
SQ0B105	SCHL0110	TOTAL EIGHTH GRADE GIFTED AND TALENTED EDUCATION
SQ00106	SCHL0111	WHAT PERCENT OF PARENTS PARTICIPATES IN PTO
SQ00107	SCHL0112	WHAT PERCENT OF PARENTS PARTICIPATES IN OPEN HOUSE
SQ00108	SCHL0113	WHAT PERCENT OF PARENTS PARTICIPATE IN CONFERENCES
SQ00109	SCHL0114	WHAT PERCENT OF PARENTS INVOLVED IN CURRICULUM DEC
SQ00110	SCHL0115	WHAT PERCENT OF PARENTS PARTICIPATES IN VOLUNTEER
SQ00111	SCHL0116	WHAT DEGREE IS STUDENT ABSENTEEISM A PROBLEM
SQ00112	SCHL0117	WHAT DEGREE IS STUDENT TARDINESS A PROBLEM
SQ00113	SCHL0118	WHAT DEGREE IS PHYSICAL CONFLICTS/CHILDREN A PROBL
SQ00114	SCHL0119	WHAT DEGREE IS TEACHER ABSENTEEISM A PROBLEM
SQ00115	SCHL0120	WHAT DEGREE IS RACIAL/CULTURAL CONFLICTS A PROBLEM
SQ00116	SCHL0121	WHAT DEGREE IS STUDENT HEALTH A PROBLEM
SQ00117	SCHL0122	WHAT DEGREE IS LACK OF PARENT INVOLVEMENT A PROBLE
SQ00118	SCHL0123	WHAT DEGREE IS STUDENT USE OF ALCOHOL A PROBLEM
SQ00119	SCHL0124	WHAT DEGREE IS STUDENT USE OF TOBACCO A PROBLEM
SQ00120	SCHL0125	WHAT DEGREE IS STUDENT USE OF DRUGS A PROBLEM
SQ00121	SCHL0126	WHAT DEGREE IS GANG ACTIVITY A PROBLEM
SQ00122	SCHL0127	WHAT DEGREE IS STUDENT MISBEHAVIOR IN CLASS A PROB
SQ00123	SCHL0128	WHAT DEGREE IS CHEATING A PROBLEM IN YOUR SCHOOL
SQ00124	SCHL0129	HOW CHARACTERIZE MORALE OF TEACHERS IN YOUR SCHOOL
SQ00125	SCHL0130	HOW CHARACTERIZE STUDENTS' ATTITUDES/ACADEM ACHIEV
SQ00126	SCHL0131	HOW CHARACTERIZE PARENTAL SUPPORT/STUDENT ACHIEVEM
SQ00127	SCHL0132	HOW CHARACTERIZE REGARD FOR SCHOOL PROPERTY
SQ00128	SCHL0133	PERCENTAGE OF STUDENTS ABSENT ON AN AVERAGE DAY
SQ00129	SCHL0134	WHAT PERCENT TEACHERS ABSENT AVERAGE DAY
SQ00130	SCHL0135	WHAT PERCENT STUDENT STILL ENROLLED AT END OF YEAR
SQ00131	SCHL0136	WHAT PERCENT THIS YEARS EIGHTH GRADERS HELD BACK
SQ00132	SCHL0137	WHAT PERCENT FULL-TIME STARTING TEACHERS LEFT
TB00001	TCHR0001	WHAT IS YOUR SEX
TB00002	TCHR0002	WHICH BEST DESCRIBES YOUR RACIAL/ETHNIC BACKGROUND
TB00003	TCHR0003	HOW MANY YEARS TAUGHT ELEM/SECONDARY SCHOOL
TB00004	TCHR0004	WHAT STATE TEACHING CERTIFICATES HELD IN FIELD
TB00005	TCHR0005	DO YOU HAVE A TEACHING CERT IN ELEM OR JR HIGH ED

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
TB00006	TCHR0006	DO YOU HAVE TEACHING CERT IN ENGLISH/LANG ARTS
TB00007	TCHR0007	DO YOU HAVE TEACHING CERT IN ARTS EDUCATION
TB00008	TCHR0008	DO YOU HAVE TEACHING CERTIFICATE IN DANCE
TB00009	TCHR0009	DO YOU HAVE A TEACHING CERTIFICATE IN MUSIC
TB00010	TCHR0010	DO YOU HAVE TEACHING CERTIFICATE IN THEATRE/DRAMA
TB00011	TCHR0011	DO YOU HAVE TEACHING CERTIFICATE IN VISUAL ARTS
TB00012	TCHR0012	DOYOU HAVE TEACHING CERTIFICATE IN OTHER/SPECIFY
TB00013	TCHR0013	WHAT IS THE HIGHEST ACADEMIC DEGREE YOU HOLD
TB00014	TCHR0014	WERE YOUR UNDERGRAD MAJORS/EDUCATION
TB00015	TCHR0015	WERE YOUR UNDERGRAD MAJORS/ELEMENTARY EDUCATION
TB00016	TCHR0016	WERE YOUR UNDERGRAD MAJORS/SECONDARY EDUCATION
TB00017	TCHR0017	WERE YOUR UNDERGRAD MAJORS/ENGLISH LANGUAGE ARTS
TB00018	TCHR0018	WERE YOUR UNDERGRAD MAJORS/ARTS EDUCATION
TB00019	TCHR0019	WERE YOUR UNDERGRAD MAJORS/DANCE OR DANCE ED
TB00020	TCHR0020	WERE YOUR UNDERGRAD MAJORS/MUSIC OR MUSIC ED
TB00021	TCHR0021	WERE YOUR UNDERGRAD MAJORS/THEATRE OR THEATRE ED
TB00022	TCHR0022	WERE YOUR UNDERGRAD MAJORS/VISUAL ARTS
TB00023	TCHR0023	WERE YOUR UNDERGRAD MAJORS/PHYSICAL EDUCATION
TB00024	TCHR0024	WERE YOUR UNDERGRAD MAJORS/SPECIAL EDUCATION
TB00025	TCHR0025	WERE YOUR UNDERGRAD MAJORS/BILINGUAL EDUCATION/ESL
TB00026	TCHR0026	WERE YOUR UNDERGRAD MAJORS/OTHER (SPECIFY)
TB00027	TCHR0027	WERE YOUR GRADUATE MAJORS/EDUCATION
TB00028	TCHR0028	WERE YOUR GRADUATE MAJORS/ELEMENTARY EDUCATION
TB00029	TCHR0029	WERE YOUR GRADUATE MAJORS/SECONDARY EDUCATION
TB00030	TCHR0030	WERE YOUR GRADUATE MAJORS/ENGLISH LANGUAGE ARTS
TB00031	TCHR0031	WERE YOUR GRADUATE MAJORS/ARTS EDUCATION
TB00032	TCHR0032	WERE YOUR GRADUATE MAJORS/DANCE OR DANCE EDUCATION
TB00033	TCHR0033	WERE YOUR GRADUATE MAJORS/MUSIC OR MUSIC EDUCATION
TB00034	TCHR0034	WERE YOUR GRADUATE MAJORS/THEATRE OR THEATRE ED
TB00035	TCHR0035	WERE YOUR GRADUATE MAJORS/VISUAL ARTS
TB00036	TCHR0036	WERE YOUR GRADUATE MAJORS/PHYSICAL EDUCATION
TB00037	TCHR0037	WERE YOUR GRADUATE MAJORS/SPECIAL EDUCATION
TB00038	TCHR0038	WERE YOUR GRADUATE MAJORS/BILINGUAL ED OR ESL
TB00039	TCHR0039	WERE YOUR GRADUATE MAJORS/ADMINISTRATION & SUPERVI
TB00040	TCHR0040	WERE YOUR GRADUATE MAJORS/CURRICULUM & INSTRUCTION
TB00041	TCHR0041	WERE YOUR GRADUATE MAJORS/OTHER (SPECIFY)
TB00042	TCHR0042	NO GRADUATE LEVEL STUDY
TB00043	TCHR0043	WERE YOUR UNDERGRAD/GRAD MINORS EDUCATION
TB00044	TCHR0044	WERE YOUR UNDERGRAD/GRAD MINORS ELEMENTARY ED
TB00045	TCHR0045	WERE YOUR UNDERGRAD/GRAD MINORS SECONDARY ED
TB00046	TCHR0046	WERE YOUR UNDERGRAD/GRAD MINORS ENGLISH/LANG ARTS
TB00047	TCHR0047	WERE YOUR UNDERGRAD MINORS ARTS EDUCATION
TB00048	TCHR0048	WERE YOUR UNDERGRAD MINORS DANCE OR DANCE ED
TB00049	TCHR0049	WERE YOUR UNDERGRAD MINORS MUSIC OR MUSIC ED
TB00050	TCHR0050	WERE YOUR UNDERGRAD MINORS THEATRE OR THEATRE ED

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
TB00051	TCHR0051	WERE YOUR UNDERGRAD MINORS VISUAL ARTS
TB00052	TCHR0052	WERE YOUR UNDERGRAD MINORS PHYSICAL EDUCATION
TB00053	TCHR0053	WERE YOUR UNDERGRAD MINORS SPECIAL EDUCATION
TB00054	TCHR0054	WERE YOUR UNDERGRAD MINORS BILINGUAL ED OR ESL
TB00055	TCHR0055	WERE YOUR UNDERGRAD MINORS ADMINISTRATION & SUPERV
TB00056	TCHR0056	WERE YOUR UNDERGRAD MINORS CURRICULUM & INSTRUCT
TB00057	TCHR0057	WERE YOUR UNDERGRAD MINORS OTHER (SPECIFY)
TB00058	TCHR0058	NO MINOR FIELD OF STUDY
TB00059	TCHR0059	TIME SPENT ON PROF DEV IN DANCE OR DANCE EDUCATION
TB00060	TCHR0060	TIME SPENT ON PROF DEV IN MUSIC OR MUSIC EDUCATION
TB00061	TCHR0061	TIME SPENT ON PROF DEV IN THEATRE OR THEATRE ED
TB00062	TCHR0062	TIME SPENT ON PROF DEV IN VISUAL ARTS OR VA ED
TB00063	TCHR0063	HOW MANY COLLEGE-UNIV. COURSES TAKEN/DANCE
TB00064	TCHR0064	HOW MANY COLLEGE-UNIV COURSES TAKEN/MUSIC
TB00065	TCHR0065	HOW MANY COLLEGE-UNIV COURSES TAKEN/THEATRE
TB00066	TCHR0066	HOW MANY COLLEGE-UNIV COURSES TAKEN/VISUAL ARTS
TB00067	TCHR0067	COURSES/PARTICIPATION IN USE OF TELECOMMUNICATIONS
TB00068	TCHR0068	COURSES/PARTICIPATION IN USE OF TECHNOLOGY
TB00069	TCHR0069	COURSES/PARTICIPATION IN COOPERATIVE GROUP INSTRUC
TB00070	TCHR0070	COURSES/PARTICIPATION IN INTERDISCIPLINARY INSTRUCT
TB00071	TCHR0071	COURSES/PARTICIPATION IN ASSESSMENT BY PORTFOLIO
TB00072	TCHR0072	COURSES/PARTICIPATION IN PERFORMANCE BASED ASSESSM
TB00073	TCHR0073	COURSES/PARTICIPATION IN TEACHING HIGHER-ORDER
TB00074	TCHR0074	COURSES/PARTICIPATION IN TEACHING STUDENTS FROM
TB00075	TCHR0075	COURSES/PARTICIPATION IN TEACHING LEPS
TB00076	TCHR0076	COURSES/PARTICIPATION IN TEACHING STUDENTS W/SPECI
TB00077	TCHR0077	COURSES/PARTICIPATION IN CLASSROOM MGT & ORG
TB00078	TCHR0078	COURSES/PARTICIPATION IN OTHER PROF ISSUES
TB00079	TCHR0079	COURSES/PARTICIPATION IN NO PROF DEV ACTIVITIES
TB00080	TCHR0080	HOW WELL SCHOOL SYSTEM PROVIDES RESOURCES/MATERIAL
TB00081	TCHR0081	IS THERE A CURRICULUM SPECIALIST IN DANCE
TB00082	TCHR0082	IS THERE A CURRICULUM SPECIALIST IN MUSIC
TB00083	TCHR0083	IS THERE A CURRICULUM SPECIALIST IN THEATRE/DRAMA
TB00084	TCHR0084	IS THERE A CURRICULUM SPECIALIST IN VISUAL ARTS
TB00085	TCHR0085	HOW MANY SCHOOL HRS DESIGNATED PREPARATION TIME/WK
TC00001	TSUB0001	TEACH OTHER SUBJECTS THAN THEATRE
TC00002	TSUB0002	IS THEATRE INSTRUCTION PART OF K-12 CURRICULUM
TC00003	TSUB0003	IS THEATRE INSTRUCTION PART OF SHORTER CURRICULUM
TC00004	TSUB0004	IS THEATRE INSTRUCTION RELATED TO LOCAL STANDARDS
TC00005	TSUB0005	IS THEATRE INSTRUCTION INTEGRATED W/ACADEMIC CURRI
TC00006	TSUB0006	IS THEATRE INSTRUCTION INTEGRATED W/ARTS CURRIC
TC00007	TSUB0007	IS THEATRE INSTRUCTION PART OF LANG ARTS CURRIC
TC00008	TSUB0008	WE DO NOT HAVE A THEATRE CURRICULUM
TC00009	TSUB0009	IS THEATRE INSTRUCTION/NONE OF THE ABOVE
TC00010	TSUB0010	MATCH THEATRE INSTRUCTION W/NATL STAND FOR ARTS ED

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
TC00011	TSUB0011	DO YOU HAVE THEATRE CURRICULUM IN 8TH GRADE
TC00012	TSUB0012	SCHOOL STAGE FACILITIES: NO STAGE
TC00013	TSUB0013	SCHOOL STAGE FACILITIES: W/MORE THAN ONE CURTAIN
TC00014	TSUB0014	SCHOOL STAGE FACILITIES: SOUND SYSTEM
TC00015	TSUB0015	SCHOOL STAGE FACILITIES: CONSTRUCTING SCENERY
TC00016	TSUB0016	SCHOOL STAGE FACILITIES: W/LIGHTING AND CONTROLS
TC00017	TSUB0017	SCHOOL STAGE FACILITIES: DRESSING ROOMS
TC00018	TSUB0018	SCHOOL STAGE FACILITIES: COSTUME STOCK
TC00019	TSUB0019	DESCRIBE YOUR ACCESS TO SCHOOL STAGE FACILITIES
TC00020	TSUB0020	PLACED IN THEATRE CLASS ACCORDING TO ABILITY
TC00021	TSUB0021	IF STUDENTS PARTICIPATE, HOW SELECTED FOR ACTIVITY
TC00022	TSUB0022	TEACHER TIME SPENT: CREATING DRAMA
TC00023	TSUB0023	TEACHER TIME SPENT: PERFORMING, ACTING, DIRECTING
TC00024	TSUB0024	TEACHER TIME SPENT: RESPONDING TO THEATRE
TC00025	TSUB0025	ASSESS STUDENT PROGRESS USING: VIDEOTAPE
TC00026	TSUB0026	ASSESS STUDENT PROGRESS USING: OBSERVATION
TC00027	TSUB0027	ASSESS STUDENT PROGRESS USING: WRITTEN TESTS
TC00028	TSUB0028	ASSESS STUDENT PROGRESS USING: PORTFOLIOS
TC00029	TSUB0029	ASSESS STUDENT PROGRESS USING: SELF-ASSESSMENT
TC00030	TSUB0030	ASSESS STUDENT PROGRESS USING: NO FORMAL ASSMT
TC00031	TSUB0031	HOW FREQUENTLY ASSESS STUDENTS IN THEATRE
TC00032	TSUB0032	WHAT KIND OF CLASS
TC00033	TSUB0033	WHERE IS CLASS HELD
TC00034	TSUB0034	ROLE IN TEACHING THEATRE
TC00035	TSUB0035	WHAT PORTION OF YEAR DOES THIS CLASS MEET
TC00036	TSUB0036	HOW OFTEN PROVIDE INSTRUCTION IN THEATRE TO CLASS
TC00037	TSUB0037	HOW OFTEN DO STUDENTS: CREATE CHARACTERS & SCENES
TC00038	TSUB0038	HOW OFTEN DO STUDENTS: READ A PLAY SILENTLY
TC00039	TSUB0039	HOW OFTEN DO STUDENTS: READ A PLAY ALOUD
TC00040	TSUB0040	HOW OFTEN DO STUDENTS: WRITE A PLAY OR A SCENE
TC00041	TSUB0041	HOW OFTEN DO STUDENTS: DO TECH ASPECTS OF THEATRE
TC00042	TSUB0042	HOW OFTEN DO STUDENTS: PERFORM FOR AN AUDIENCE
TC00043	TSUB0043	HOW OFTEN DO STUDENTS: READ ABOUT THEATRE
TC00044	TSUB0044	HOW OFTEN DO STUDENTS: WATCH TV, VIDEO- THEATRE
TC00045	TSUB0045	HOW OFTEN DO STUDENTS: DO RESEARCH PROJECT-THEATRE
TC00046	TSUB0046	HOW OFTEN DO STUDENTS: DIRECT A PLAY OR SCENE
TC00047	TSUB0047	HOW OFTEN DO STUDENTS: CRITIQUE A PLAY
TC00048	TSUB0048	TIME SPENT DOING THEATRE HOMEWORK OR PRACTICE/WK
TC00049	TSUB0049	FIELD TRIPS TO SEE THEATRE PRODUCTIONS/YR
TP00001	TCHR0086	PARTICIPATE IN PROF DEV ACTIVITY/CREATIVE DRAMA
TP00002	TCHR0087	PARTICIPATE IN PROF DEV ACTIVITY/THEATRE IN ED
TP00003	TCHR0088	PARTICIPATE PROF ACTIV/METHOD TEACHING THEATRE ED
TP00004	TCHR0089	PARTICIPATE PROF ACTIVITY/FORENSICS
TP00005	TCHR0090	PARTICIPATE PROF DEV ACTIVITY/PERFORMANCE STUDIES
TP00006	TCHR0091	PARTICIPATE IN PROF DEV ACTIVITY/DRAMA THERAPY

(continued)

Table G-3 (continued)
Summary Table of the 1997 Theatre Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
TP00007	TCHR0092	PARTICIPATE IN PROF DEV ACTIVITY/ARTS ADMIN
TP00008	TCHR0093	PARTICIPATE IN PROF DEV ACTIVITY/FILMMAKING
TP00009	TCHR0094	PARTICIPATE IN PROF DEV ACTIVITY/PLAY PRODUCTION
TP00010	TCHR0095	PARTICIPATE IN PROF DEV ACTIVITY/TECHNICAL THEATRE
TP00011	TCHR0096	PARTICIPATE IN PROF DEV ACTIVITY/DESIGN
TP00012	TCHR0097	PARTICIPATE IN PROF DEV ACTIVITY/PLAY-SCREENWRITIN
TP00013	TCHR0098	PARTICIPATE IN PROF DEV ACTIVITY/ACTING
TP00014	TCHR0099	PARTICIPATE IN PROF DEV ACTIVITY/DIRECTING
TP00015	TCHR0100	PARTICIPATE IN PROF DEV ACTIVITY/OTHER (SPECIFY)
TP00016	TCHR0101	PARTICIPATE IN PROF DEV ACTIVITY/NONE OF THE ABOVE
TP00017	TCHR0102	HOW OFTEN ATTEND LIVE THEATRE PRODUCTIONS
TP00018	TCHR0103	PARTICIPATE THEATRE RELATED ACTIVITY/ACTING PROFES
TP00019	TCHR0104	PARTICIPATE THEATRE RELATED ACTIVITY/DIRECTING
TP00020	TCHR0105	PARTICIPATE THEATRE RELATED ACTIVITY/ AVOCATION
TP00021	TCHR0106	PARTICIPATE THEATRE RELATED ACTIVITY/WRITE PLAYS
TP00022	TCHR0107	PARTICIPATE THEATRE RELATED ACTIVITY/JUDGING COMPE
TP00023	TCHR0108	PARTICIPATE THEATRE RELATED ACTIVITY/WRITE CRITIC
TP00024	TCHR0109	PARTICIPATE THEATRE RELATED ACTIVITY/OTHER
TP00025	TCHR0110	PARTICIPATE THEATRE RELATED ACTIVITY/NONE

Table G-4
Summary Table of the 1997 Visual Arts Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
BKSER	BACK0001	GRAND MEAN
DSEX	BACK0002	DERIVED SEX
DRACE	BACK0003	DERIVED RACE
NAEPRGN	BACK0004	NAEP REGION (1:NE;2:SE;3:CENT;4:WEST)
SCHTYP7	BACK0005	SCHOOL TYPE
TOL8	BACK0006	MSA/NON-MSA
TOL5	BACK0007	TYPE OF LOCALE (5 CATEGORIES)
DOC	BACK0008	DESCRIPTION OF COMMUNITY
PARED	BACK0009	PARENTS' HIGHEST LEVEL OF EDUCATION
HOMEEN	BACK0010	NUMBER OF ITEMS IN THE HOME (NEWSPAPER, > 25 BOOKS, ENCYCLOPEDIA, MAGAZINES)(DERIVED)
IEP	BACK0011	INDIVIDUALIZED EDUCATION PLAN
LEP	BACK0012	LIMITED ENGLISH PROFICIENCY
MUSELIG	BACK0013	ELIGIBLE FOR MUSIC
MUSADV	BACK0014	MUSIC EDVANCED
B000002	BACK0015	ARE YOU HISPANIC
B000003	BACK0016	HOW LONG HAVE YOU LIVED IN UNITED STATES
B000004	BACK0017	HOW OFTEN ENGLISH SPOKEN AT HOME
B000005	BACK0018	HOW FAR IN SCHOOL DID YOUR MOTHER GO

(continued)

Table G-4 (continued)
Summary Table of the 1997 Visual Arts Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
B000006	BACK0019	HOW FAR IN SCHOOL DID YOUR FATHER GO
B000007	BACK0020	DOES YOUR FAMILY GET A NEWSPAPER REGULARLY
B000008	BACK0021	IS THERE AN ENCYCLOPEDIA IN YOUR HOME
B000009	BACK0022	HOW MANY BOOKS IN YOUR HOME?
B000010	BACK0023	DOES YOUR FAMILY GET MAGAZINES REGULARLY
B000011	BACK0024	HOW MUCH TELEVISION DO YOU USUALLY WATCH EACH DAY
B000012	BACK0025	HOW MUCH TIME SPENT OF HOMEWORK EACH DAY
B000013	BACK0026	HOW OFTEN DO YOU READ FOR FUN ON OWN TIME
B000014	BACK0027	HOW MANY DAYS OF SCHOOL DID YOU MISS LAST MONTH
B000015	BACK0028	DID YOU GO TO PRESCHOOL BEFORE KINDERGARTEN
B000016	BACK0029	HOW MANY TIMES IN 2 YRS CHANGED SCHOOLS - MOVED
B000017	BACK0030	WITHIN PAST 2 YRS DID YOU CHANGE SCHOOLS - NO MOVE
B000018	BACK0031	HOW OFTEN DO YOU DISCUSS STUDIES IN SCHOOL AT HOME
B000019	BACK0032	HOW OFTEN DO YOU USE A COMPUTER AT HOME FOR SCHOOL
B000020	BACK0033	READ IN SCHOOL & HOMEWORK
B000021	BACK0034	WHICH STATEMENT BEST DESCRIBES GRADES SINCE GR 6
B000022	BACK0035	HOW MUCH EDUCATION DO YOU EXPECT TO RECEIVE
BV00001	SUBJ0001	I LIKE TO LOOK AT ART
BV00002	SUBJ0002	I LIKE TO DO ART WORK
BV00003	SUBJ0003	I THINK I HAVE A TALENT FOR ART
BV00004	SUBJ0004	PEOPLE TELL ME I AM A GOOD ARTIST
BV00005	SUBJ0005	I LIKE TO SHOW MY ARTWORK TO OTHER PEOPLE
BV00006	SUBJ0006	I WOULD LIKE TO BE AN ARTIST WHEN I GROW UP
BV00007	SUBJ0007	ARE YOU TAKING AN ART COURSE NOW, OR HAVE THIS YR
BV00008	SUBJ0008	HOW OFTEN DOES TEACHER HAVE YOU PAINT OR DRAW
BV00009	SUBJ0009	MAKE THINGS OUT OF CLAY OR OTHER MATERIAL
BV00010	SUBJ0010	CHOOSE YOUR OWN PROJECT
BV00011	SUBJ0011	WORK IN A PAIR OR A GROUP ON AN ART PROJECT
BV00012	SUBJ0012	TALK W/OTHERS ABOUT YOUR ARTWORK OR OTHER STUDENTS
BV00013	SUBJ0013	WRITE ABOUT YOUR ARTWORK
BV00014	SUBJ0014	LOOK AT VIDEOS, FILMS, SLIDES, OR TV ABOUT ART
BV00015	SUBJ0015	WORK WITH A CAMERA, COMPUTER, XEROX TO MAKE ART
BV00016	SUBJ0016	HOW OFTEN DOES YOUR TEACHER EXHIBIT YOUR ARTWORK
BV00017	SUBJ0017	DO YOU EVER ILLUSTRATE YOUR WORK IN OTHER SUBJECTS
BV00018	SUBJ0018	DO YOU KEEP AN ART JOURNAL OR SKETCHBOOK IN SCHOOL
BV00019	SUBJ0019	DO YOU OR YOUR TEACHER SAVE ARTWORK IN PORTFOLIO
BV00020	SUBJ0020	HOW OFTEN DO YOU HAVE HOMEWORK FOR ART CLASS
BV00021	SUBJ0021	HOW MANY TIMES DID YOUR CLASS VISIT ART MUSEUM ETC
BV00022	SUBJ0022	GO TO AN ART MUSEUM OR EXHIBIT
BV00023	SUBJ0023	TAKE ART CLASS
BV00024	SUBJ0024	MAKE ARTWORK
BV00025	SUBJ0025	EXHIBIT YOUR ARTWORK
BV00026	SUBJ0026	ENTER AN ART COMPETITION
BV00027	SUBJ0027	GO TO A SUMMER ART PROGRAM
BV00028	SUBJ0028	LOOK AT OR READ A BOOK ABOUT ART

(continued)

Table G-4 (continued)
Summary Table of the 1997 Visual Arts Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
BV00029	SUBJ0029	WATCH A VIDEO OR TV PROGRAM ABOUT ART
BV00030	SUBJ0030	TALK WITH FAMILY OR FRIENDS ABOUT ART
BV00031	SUBJ0031	VISIT AN ARTIST'S STUDIO
BV00032	SUBJ0032	KEEP AN ART JOURNAL OR SKETCHBOOK
BV00033	SUBJ0033	NONE OF THE ABOVE
SQ00004	SCHL0001	VISUAL ARTS
SQ00005	SCHL0002	DANCE
SQ00006	SCHL0003	MUSIC
SQ00007	SCHL0004	THEATRE
SQ00008	SCHL0005	VISUAL ARTS
SQ00009	SCHL0006	COMPUTERS AVAILABLE IN CLASSROOM
SQ00010	SCHL0007	COMPUTERS IN LABORATORIES
SQ00011	SCHL0008	BRING COMPUTERS TO CLASSROOM
SQ00012	SCHL0009	SPECIAL FOCUS DANCE
SQ00013	SCHL0010	SPECIAL FOCUS MUSIC
SQ00014	SCHL0011	SPECIAL FOCUS THEATRE
SQ00015	SCHL0012	SPECIAL FOCUS VISUAL ARTS
SQ00016	SCHL0013	SPECIAL FOCUS OTHER
SQ00017	SCHL0014	NOT A SPECIAL FOCUS SCHOOL
SQ00018	SCHL0015	DANCE CURRICULUM
SQ00019	SCHL0016	MUSIC CURRICULUM
SQ00020	SCHL0017	THEATRE CURRICULUM
SQ00021	SCHL0018	CURRICULUM VISUAL ARTS
SQ00022	SCHL0019	CURRICULUM NONE OF THE ABOVE
SQ00023	SCHL0020	DANCE FIELD TRIPS
SQ00024	SCHL0021	MUSIC FIELD TRIPS
SQ00025	SCHL0022	THEATRE FIELD TRIPS
SQ00026	SCHL0023	VISUAL ARTS FIELD TRIPS
SQ00027	SCHL0024	OTHER (SPECIFY) FIELD TRIPS
SQ00028	SCHL0025	NONE OF THE ABOVE FIELD TRIPS
SQ00029	SCHL0026	EXTRACURRICULAR SPONSORED DANCE
SQ00030	SCHL0027	EXTRACURRICULAR SPONSORED MUSIC
SQ00031	SCHL0028	EXTRACURRICULAR SPONSORED THEATRE
SQ00032	SCHL0029	EXTRACURRICULAR SPONSORED VISUAL ARTS
SQ00033	SCHL0030	EXTRACURRICULAR SPONSORED NONE OF THE ABOVE
SQ00034	SCHL0031	SUMMER PROGRAM DANCE
SQ00035	SCHL0032	SUMMER PROGRAM MUSIC
SQ00036	SCHL0033	SUMMER PROGRAM THEATRE
SQ00037	SCHL0034	SUMMER PROGRAM VISUAL ARTS
SQ00038	SCHL0035	SUMMER PROGRAMS NONE OF THE ABOVE
SQ00039	SCHL0036	POSITION OF STAFF - DANCE FULL-TIME SPECIALIST
SQ00040	SCHL0037	POSITION OF STAFF - DANCE PART-TIME SPECIALIST
SQ00041	SCHL0038	POSITION OF STAFF- DANCE PHYSICAL ED TEACHER
SQ00042	SCHL0039	POSITION OF STAFF- DANCE ELEMENTARY TEACHER
SQ00043	SCHL0040	POSITION OF STAFF-OTHER FACULTY MEMBER

(continued)

Table G-4 (continued)
Summary Table of the 1997 Visual Arts Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
SQ00044	SCHL0041	POSITION OF STAFF- DANCE ARTIST IN RESIDENCE
SQ00045	SCHL0042	POSITION OF STAFF- DANCE VOLUNTEER
SQ00046	SCHL0043	POSITION OF STAFF-DANCE OR MOVEMENT NOT TAUGHT
SQ00047	SCHL0044	POSITION OF STAFF-MUSIC FULLTIME SPECIALIST
SQ00048	SCHL0045	POSITION OF STAFF-MUSIC PART TIME SPECIALIST
SQ00049	SCHL0046	POSITION OF STAFF-MUSIC ELEMENTARY TEACHER
SQ00050	SCHL0047	POSITION OF STAFF-MUSIC OTHER FACULTY MEMBER
SQ00051	SCHL0048	POSITION OF STAFF-MUSIC ARTIST IN RESIDENCE
SQ00052	SCHL0049	POSITIONOF STAFF-MUSIC VOLUNTEER
SQ00053	SCHL0050	POSITION OF STAFF-MUSIC NOT TAUGHT
SQ00054	SCHL0051	POSITION OF STAFF -THEATRE FULL TIME SPECIALIST
SQ00055	SCHL0052	POSITION OF STAFF-THEATRE PART TIME SPECIALIST
SQ00056	SCHL0053	POSITION OF STAFF-THEATRE ENGLISH LANG ARTS TEACHE
SQ00057	SCHL0054	POSITION OF STAFF-THEATRE ELEMENTARY TEACHER
SQ00058	SCHL0055	POSITION OF STAFF-THEATRE OTHER FACULTY MEMBER
SQ00059	SCHL0056	POSITION OF STAFF-THEATRE ARTIST IN RESIDENCE
SQ00060	SCHL0057	POSITION OF STAFF-THEATRE VOLUNTEER
SQ00061	SCHL0058	POSITION OF STAFF-THEATRE OR DRAMA NOT TAUGHT
SQ00062	SCHL0059	POSITION OF STAFF-VISUAL ARTS FULLTIME SPECIALIST
SQ00063	SCHL0060	POSITION OF STAFF-VISUAL ARTS PART TIME SPECIALIST
SQ00064	SCHL0061	POSITION OF STAFF-VISUAL ARTS ELEMENTARY TEACHER
SQ00065	SCHL0062	POSITION OF STAFF-VISUAL ARTS OTHER FACULTY MEMBER
SQ00066	SCHL0063	POSITIONOF STAFF-VISUAL ARTS ARTIST IN RESIDENCE
SQ00067	SCHL0064	POSITION OF STAFF-VISUAL ARTS VOLUNTEER
SQ00068	SCHL0065	POSITION OF STAFF-VISUAL ARTS NOT TAUGHT
SQ00069	SCHL0066	WHICH BEST DESCRIBES SPACE FOR TEACHING DANCE
SQ00070	SCHL0067	WHICH BEST DESCRIBES SPACE FOR PERFORMING MUSIC
SQ00071	SCHL0068	WHICH BEST DESCRIBES SPACE FOR TEACHING THEATRE
SQ00072	SCHL0069	WHICH BEST DESCRIBES SPACE FOR TEACHING VISUAL ART
SQ00073	SCHL0070	DID SCHOOL BRING VISITING ARTISTS - DANCE
SQ00074	SCHL0071	DID SCHOOL BRING VISITING ARTISTS - MUSIC
SQ00075	SCHL0072	DID SCHOOL BRING VISITING ARTISTS - THEATRE
SQ00076	SCHL0073	DID SCHOOL BRING VISITING ARTISTS - VISUAL ARTS
SQ00077	SCHL0074	DID SCHOOL BRING VISTING ARTISTS - NONE
SQ00078	SCHL0075	DID SCHOOL SPONSOR VISITING ARTIST PROG -DANCE
SQ00079	SCHL0076	DID SCHOOL SPONSOR VISITING ARTIST PROG - FILM
SQ00080	SCHL0077	DID SCHOOL SPONSOR VISITING ARTIST PROG-FOLK ARTS
SQ00081	SCHL0078	DID SCHOOL SPONSOR VISITING ARTIST PROG-MUSIC
SQ00082	SCHL0079	DID SCHOOL SPONSOR VISITING ARTISTS PROG-THEATRE
SQ00083	SCHL0080	DID SCHOOL SPONSOR VISITING ARTIST PROG-VISUAL ART
SQ00084	SCHL0081	DID SCHOOL SPONSOR VISITING ARTIST PROG-NONE
SQ00085	SCHL0082	WHAT PERCENT 8TH GRADE INSTRUCTED - DANCE
SQ00086	SCHL0083	WHAT PERCENT 8TH GRADE INSTRUCTED - MUSIC
SQ00087	SCHL0084	WHAT PERCENT 8TH GRADE INSTRUCTED - THEATRE
SQ00088	SCHL0085	WHAT PERCENT 8TH GRADE INSTRUCTED - VISUAL ARTS

(continued)

Table G-4 (continued)
Summary Table of the 1997 Visual Arts Variables Defining Group Membership

Grouping Variable ID	NAEP ID	Description
SQ00089	SCHL0086	WHAT IS THE CURRENT ENROLLMENT IN YOUR SCHOOL?
SQ00090	SCHL0087	WHAT IS THE CURRENT ENROLLMENT IN THE EIGHTH GRADE
SQ00091	SCHL0088	DOES SCHOOL PARTICIPATE-NATIONAL SCHOOL LUNCH PROG
SQ00092	SCHL0089	TOTAL # STUDENTS ELIGIBLE-NATL SCHOOL LUNCH PROG
SQ00093	SCHL0090	HOW MANY 8TH GRADE ELIGIBLE NATL SCHOOL LUNCH PROG
SQ00094	SCHL0091	DOES SCHOOL RECEIVE CHPATER1/TITLE 1 FUNDING
SQ00095	SCHL0092	TOTAL # STUDENTS ELIGIBLE CHAPTER 1/TITLE 1
SQ00096	SCHL0093	HOW MANY 8TH GRADE ELIGIBLE CHAPTER 1/TITLE 1
SQ00097	SCHL0094	WHAT PERCENT STUDENTS ELIGIBLE CHAPTER1/TITLE 1
SQ00A98	SCHL0095	TOTAL SCHOOL NATIONAL SCHOOL LUNCH PROGRAM
SQ00B98	SCHL0096	TOTAL EIGHTH GRADE NATIONAL SCHOOL LUNCH PROGRAM
SQ00A99	SCHL0097	TOTAL SCHOOL CHAPTER 1/TITLE 1
SQ00B99	SCHL0098	TOTAL EIGHTH GRADE CHAPTER1/TITLE 1
SQ0A100	SCHL0099	TOTAL SCHOOL REMEDIAL READING INSTRUCTION
SQ0B100	SCHL0100	TOTAL EIGHTH GRADE REMEDIAL READING INSTRUCTION
SQ0A101	SCHL0101	TOTAL SCHOOL REMEDIAL MATH INSTRUCTION
SQ0B101	SCHL0102	TOTAL EIGHTH GRADE REMEDIAL MATH INSTRUCTION
SQ0A102	SCHL0103	TOTAL SCHOOL BILINGUAL EDUCATION
SQ0B102	SCHL0104	TOTAL EIGHTH GRADE BILINGUAL EDUCATION
SQ0A103	SCHL0105	TOTAL SCHOOL ENGLISH AS SECOND LANGUAGE INSTRUCT
SQ0B103	SCHL0106	EIGHTH GRADE ENGLISH AS SECOND LANGUAGE INSTRUCT
SQ0A104	SCHL0107	TOTAL SCHOOL SPECIAL EDUCATION PARTICIPANTS
SQ0B104	SCHL0108	TOTAL EIGHTH GRADE SPECIAL EDUCATION PARTICIPANTS
SQ0A105	SCHL0109	TOTAL SCHOOL GIFTED AND TALENTED EDUCATION PARTICI
SQ0B105	SCHL0110	TOTAL EIGHTH GRADE GIFTED AND TALENTED EDUCATION
SQ00106	SCHL0111	WHAT PERCENT OF PARENTS PARTICIPATES IN PTO
SQ00107	SCHL0112	WHAT PERCENT OF PARENTS PARTICIPATES IN OPEN HOUSE
SQ00108	SCHL0113	WHAT PERCENT OF PARENTS PARTICIPATE IN CONFERENCES
SQ00109	SCHL0114	WHAT PERCENT OF PARENTS INVOLVED IN CURRICULUM DEC
SQ00110	SCHL0115	WHAT PERCENT OF PARENTS PARTICIPATES IN VOLUNTEER
SQ00111	SCHL0116	WHAT DEGREE IS STUDENT ABSENTEEISM A PROBLEM
SQ00112	SCHL0117	WHAT DEGREE IS STUDENT TARDINESS A PROBLEM
SQ00113	SCHL0118	WHAT DEGREE IS PHYSICAL CONFLICTS/CHILDREN A PROBL
SQ00114	SCHL0119	WHAT DEGREE IS TEACHER ABSENTEEISM A PROBLEM
SQ00115	SCHL0120	WHAT DEGREE IS RACIAL/CULTURAL CONFLICTS A PROBLEM
SQ00116	SCHL0121	WHAT DEGREE IS STUDENT HEALTH A PROBLEM
SQ00117	SCHL0122	WHAT DEGREE IS LACK OF PARENT INVOLVEMENT A PROBLE
SQ00118	SCHL0123	WHAT DEGREE IS STUDENT USE OF ALCOHOL A PROBLEM
SQ00119	SCHL0124	WHAT DEGREE IS STUDENT USE OF TOBACCO A PROBLEM
SQ00120	SCHL0125	WHAT DEGREE IS STUDENT USE OF DRUGS A PROBLEM
SQ00121	SCHL0126	WHAT DEGREE IS GANG ACTIVITY A PROBLEM
SQ00122	SCHL0127	WHAT DEGREE IS STUDENT MISBEHAVIOR IN CLASS A PROB
SQ00123	SCHL0128	WHAT DEGREE IS CHEATING A PROBLEM IN YOUR SCHOOL
SQ00124	SCHL0129	HOW CHARACTERIZE MORALE OF TEACHERS IN YOUR SCHOOL
SQ00125	SCHL0130	HOW CHARACTERIZE STUDENTS' ATTITUDES/ACADEM ACHIEV

(continued)

Table G-4 (continued)*Summary Table of the 1997 Visual Arts Variables Defining Group Membership*

Grouping Variable ID	NAEP ID	Description
SQ00126	SCHL0131	HOW CHARACTERIZE PARENTAL SUPPORT/STUDENT ACHIEVEM
SQ00127	SCHL0132	HOW CHARACTERIZE REGARD FOR SCHOOL PROPERTY
SQ00128	SCHL0133	PERCENTAGE OF STUDENTS ABSENT ON AN AVERAGE DAY
SQ00129	SCHL0134	WHAT PERCENT TEACHERS ABSENT AVERAGE DAY
SQ00130	SCHL0135	WHAT PERCENT STUDENT STILL ENROLLED AT END OF YEAR
SQ00131	SCHL0136	WHAT PERCENT THIS YEARS EIGHTH GRADERS HELD BACK
SQ00132	SCHL0137	WHAT PERCENT FULL-TIME STARTING TEACHERS LEFT

Reference List

Allen, N. L., Carlson, J. E., & Zelenak, C. Z. (1999). *The NAEP 1996 technical report* (NCES 199-452). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 technical report* (NCES 2001-509). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, DC: Author.

Calderone, J., King, L. M., & Horkay, N. (Eds.). (1997). *The NAEP guide: A description of the content and methods of the 1997 and 1998 assessments* (NCES 97-990). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.

Educational Testing Service. (2003). *ETS standards for quality and fairness*. Princeton, NJ: Author.

Mazzeo, J., Johnson, E. G., Bowker, D., & Fong, Y. F. (1992). *The use of collateral information in proficiency estimation for the trial state assessment*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mislevy, R. J., & Bock, R. D. (1982). *BILOG: Item analysis and test scoring with binary logistic models* [Computer program]. Mooresville, IN: Scientific Software.

Miller, R.G. (1996). *Simultaneous statistical inference*. New York: Wiley.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16(2), 159–176.

Muraki, E., & Bock, R. D. (1991). *PARSCALE: Parameter scaling of rating data*. Chicago, IL: Scientific Software, Inc.

National Assessment Governing Board. (1994). *NAEP arts education consensus project: 1997 arts education assessment framework*. Washington, DC: Author.

National Center for Education Statistics. (2003). *NCES statistical standards*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Persky, H. R., Sandene, B. A., & Askew, J. M. (1998). *The NAEP 1997 arts report card* (NCES 1999-486). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Rogers, A. M., Yan, F., Schoeps, T. L., & Kline, D. L. (2000). *The NAEP 1997 arts data companion*. Princeton, NJ: Educational Testing Service.

Vanneman, A., & Goodwin, M. (1998). *Focus on NAEP: NAEP and visual arts: Framework, field tests, and assessment* (NCES 98-526). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Vanneman, A., Morton, C., & Allen, L. B. (1998). *Focus on NAEP: NAEP and theatre: Framework, field tests, and assessment* (NCES 98-528). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Vanneman, A., Schuler, S., & Sandene, B. (1998). *Focus on NAEP: NAEP and music: Framework, field tests, and assessment* (NCES 98-529). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

White, S., & Vanneman, A. (1998a). *Focus on NAEP: The NAEP 1997 arts education assessment: an overview* (NCES 98-527). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

White, S., & Vanneman, A. (1998b). *Focus on NAEP: NAEP and dance: Framework and field tests* (NCES 98-459). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.