



Alaska Alternate Assessment

2011 - 2012

Technical Report

Dillard Research Associates
June 30, 2012

Alaska Alternate Assessment Technical Report

Table of Contents

| | |
|---|-----------|
| EXECUTIVE SUMMARY | 1 |
| CHAPTER 1: BACKGROUND OF THE ALASKA ALTERNATE ASSESSMENT | 3 |
| Overview | 3 |
| <i>History of Previous Program</i> | 3 |
| Current Program Overview | 3 |
| <i>Reasons for Current Approach</i> | 3 |
| <i>Roles of Contractor, Department, and Others</i> | 3 |
| Summary of Current Program | 4 |
| <i>Description of Program</i> | 4 |
| <i>Description of Students Served</i> | 4 |
| <i>Description of How Scores Are Used</i> | 5 |
| Significant Changes Since Previous Technical Report | 5 |
| Organization of Technical Report | 6 |
| CHAPTER 2: TEST DESIGN AND ITEM/TASK DEVELOPMENT | 7 |
| Overview | 7 |
| <i>Description of ExGLEs and their relationship to GLEs</i> | 8 |
| Test Specifications and Blueprint | 8 |
| <i>Description of Test Specifications (DOTS)</i> | 8 |
| <i>Process of Establishing Test Specifications</i> | 9 |
| <i>Item Content Test Blueprint and Item Specifications</i> | 10 |
| <i>Proficiency Level Descriptor Development</i> | 10 |
| <i>Cut Scores</i> | 11 |
| Item/Task Development | 12 |
| <i>Item Writing, including Scoring Guides</i> | 12 |
| <i>Item Development and Expanded Levels of Support (ELOS)</i> | 12 |
| <i>Reduction in Complexity, Depth, and Breadth</i> | 17 |
| <i>Bias and Sensitivity Review</i> | 18 |
| Test Design and Development | 19 |
| <i>Representation and Functionality</i> | 19 |
| <i>Psychometric Guidelines for Selecting Items/Tasks for Item/Task Bank</i> | 19 |
| <i>Item Bank Summary</i> | 19 |
| Current Test Construction | 20 |
| <i>Construction of the Operational Forms</i> | 20 |
| <i>Test Development Timeline</i> | 20 |
| Possible Future Test Construction | 21 |
| <i>Psychometric Guidelines for Constructing Future Forms</i> | 21 |
| <i>Possible Psychometric Guidelines for Constructing Future Forms</i> | 21 |
| <i>Internal Review of the Items and Forms</i> | 25 |
| CHAPTER 3: TEST ADMINISTRATION PROCEDURES | 26 |
| Overview | 26 |
| Student Population Tested | 26 |
| <i>Standard Administration With or Without Accommodations</i> | 26 |

| | |
|---|-----------|
| <i>Standard Administration With or Without Accommodations AND Then Switched to the ELOS</i> | 27 |
| <i>ELOS Administration</i> | 28 |
| Accommodations | 28 |
| Test Administrators | 29 |
| Mentor Responsibilities | 29 |
| Materials | 30 |
| Test Administrator Training | 30 |
| <i>New Mentor Training</i> | 31 |
| <i>Summary of Dates and Participants</i> | 31 |
| <i>Scoring Accuracy Analysis</i> | 31 |
| <i>Annual Mentor Training</i> | 32 |
| <i>Webinar</i> | 32 |
| <i>Online Training</i> | 33 |
| <i>Refresher Training and Testing</i> | 33 |
| <i>Security</i> | 33 |
| CHAPTER 4: SCORING | 35 |
| Overview | 35 |
| Quality Control of Scoring | 35 |
| <i>Procedures</i> | 35 |
| <i>Inter-rater Agreement</i> | 36 |
| <i>Handling of Exceptional Cases</i> | 36 |
| <i>Writing Study</i> | 36 |
| Data Entry | 37 |
| <i>Analysis of Accommodations Used</i> | 40 |
| CHAPTER 5: STANDARDS VALIDATION | 42 |
| CHAPTER 6: REPORTING | 43 |
| Overview | 43 |
| Reporting Student Results | 44 |
| DRA Secure Reporting Website | 44 |
| CHAPTER 7: TEST VALIDITY | 45 |
| Overview | 45 |
| Validity | 45 |
| Reliability | 46 |
| Total Test Reliabilities (All Students) | 47 |
| Test Task Reliabilities | 48 |
| <i>Reading Reliability</i> | 48 |
| <i>Writing Reliability</i> | 48 |
| <i>Math Reliability</i> | 49 |
| <i>Science Reliability</i> | 50 |
| CHAPTER 8: DESCRIPTIVE STATISTICS | 51 |
| Strand, Task, and Item Difficulties | 51 |
| <i>Strand Difficulties in Reading, Writing, Math, and Science (Standard, No ELOS)</i> | 51 |
| <i>Reading Strand Difficulties</i> | 51 |
| <i>Writing Strand Difficulties</i> | 52 |
| <i>Mathematics Strand Difficulties</i> | 52 |
| <i>Science Strand Difficulties</i> | 53 |
| <i>Task Difficulty (Standard, No ELOS)</i> | 54 |

| | |
|--|-----------|
| <i>Reading Task Difficulties</i> | 54 |
| <i>Writing Task Difficulties</i> | 55 |
| <i>Mathematics Task Difficulties</i> | 57 |
| <i>Science Task Difficulties</i> | 59 |
| <i>Item Difficulty Analyses (Standard, No ELOS)</i> | 60 |
| <i>Reading Item Difficulties</i> | 60 |
| <i>Writing Item Difficulties</i> | 66 |
| <i>Mathematics Item Difficulties</i> | 71 |
| <i>Science Item Difficulties</i> | 78 |
| <i>ELOS Task Level of Support Analysis</i> | 80 |
| <i>ELOS Item Order Analysis</i> | 88 |
| CHAPTER 9: ADEQUATE YEARLY PROGRESS | 91 |
| <i>Standard</i> | 91 |
| <i>Individual Student Reports</i> | 91 |
| <i>Confirmatory Quality Assurance Review</i> | 92 |
| <i>Adequate Yearly Progress Report Overview</i> | 92 |
| <i>Reading Adequate Yearly Progress</i> | 92 |
| <i>Writing Adequate Yearly Progress</i> | 93 |
| <i>Math Adequate Yearly Progress</i> | 93 |
| <i>Science Adequate Yearly Progress</i> | 94 |
| CHAPTER 10: PROGRAM IMPROVEMENT | 95 |
| Program Evaluation | 95 |
| Summary of Consequential Survey | 95 |
| <i>Training and Qualifications</i> | 95 |
| <i>Instructional Relevance</i> | 95 |
| <i>Revised ELOS Documents & Assessment</i> | 95 |
| <i>Revised ELOS Administration Format and Procedures</i> | 95 |
| <i>Teacher Demographics and Experiences</i> | 96 |
| Recommendations for Future Consideration | 96 |
| <i>Technological Improvements</i> | 96 |
| Recommendations for Training | 96 |
| <i>Training Recommendations from EED-led trainings</i> | 97 |
| <i>Training Recommendations from Accommodations Used</i> | 97 |
| <i>Training Recommendations from the Writing Study</i> | 97 |
| <i>Training Recommendations from Proficiency Testing</i> | 98 |
| <i>Training Recommendations from HelpDesk Questions</i> | 98 |

APPENDICES

| | |
|---|----|
| <i>Appendix 1.1 Web Changes Handout</i> | 6 |
| <i>Appendix 1.2 ELOS Item Maps 2012</i> | 6 |
| <i>Appendix 2.1 Strands: Reading, Writing, Mathematics, and Science</i> | 10 |
| <i>Appendix 2.2 Weights: Reading, Writing, and Mathematics</i> | 10 |
| <i>Appendix 2.3 Proficiency Level Descriptors</i> | 10 |
| <i>Appendix 2.4 Writing Training Documents</i> | 12 |
| <i>Appendix 2.5 Instrument 1 - Linguistic Complexity & Instrument 2 – Bias and Sensitivity Review</i> | 18 |
| <i>Appendix 2.6 Equating Map</i> | 23 |
| <i>Appendix 2.7 ACAA 2010-11 Quality Assurance</i> | 25 |
| <i>Appendix 3.1a New Mentor Training Attendees</i> | 31 |
| <i>Appendix 3.1b New Mentor Training Agenda</i> | 31 |
| <i>Appendix 3.1c New Mentor Training Handouts</i> | 31 |
| <i>Appendix 3.2a Annual Mentor Training Attendees</i> | 32 |
| <i>Appendix 3.2b Annual Mentor Training Agenda</i> | 32 |
| <i>Appendix 3.2c Annual Mentor Training Handouts</i> | 32 |
| <i>Appendix 3.3 Webinar</i> | 32 |
| <i>Appendix 3.4 Training Site Table of Contents</i> | 33 |
| <i>Appendix 3.5 Website Report Specifications</i> | 33 |
| <i>Appendix 3.6 Refresher Training Tasks</i> | 33 |
| <i>Appendix 3.7 Test Security Agreement</i> | 34 |
| <i>Appendix 3.8 Test Site Security</i> | 34 |
| <i>Appendix 4.1 Proficiency Test Item Analysis</i> | 35 |
| <i>Appendix 4.2 New Mentor Inter-rater Reliability Report</i> | 36 |
| <i>Appendix 4.3a-4.3m Practice Tests</i> | 36 |
| <i>Appendix 4.4 Writing Scoring Study Documents</i> | 37 |
| <i>Appendix 4.5 Accommodations Used Summary</i> | 40 |
| <i>Appendix 6.1 Unofficial and Official Individual Student Report Matrix</i> | 44 |
| <i>Appendix 6.2a Educator Reading, Writing, Math and Science Report Examples</i> | 44 |
| <i>Appendix 6.2b Parent Reading, Writing, Math and Science Report Examples</i> | 44 |
| <i>Appendix 6.3 Reporting Website Manual</i> | 44 |
| <i>Appendices 7.1 – 7.4 Reading, Writing, Math, and Science Reliability Statistics</i> | 47 |
| <i>Appendix 7.1 Reading Statistics</i> | 93 |
| <i>Appendix 7.2 Writing Statistics</i> | 93 |
| <i>Appendix 7.3 Mathematics Statistics</i> | 94 |
| <i>Appendix 7.4 Science Statistics</i> | 94 |
| <i>Appendix 10.1a Consequential Validity Survey</i> | 96 |
| <i>Appendix 10.1b Consequential Validity Survey Summary</i> | 96 |
| <i>Appendix 10.1c Consequential Validity Survey Written Responses</i> | 96 |
| <i>Appendix 10.2 HelpDesk Log</i> | 98 |

EXECUTIVE SUMMARY

The purpose of this Technical Report is to record the administration and reporting of the 2011-2012 Alaska Alternate Assessment.

The critical elements highlighted in this document, with examples of acceptable evidence, include (a) academic content standards, (b) academic achievement standards, (c) a statewide assessment system, (d) validity, (e) reliability, and (f) other dimensions of technical quality. We address the critical elements of training, administration, scoring, and reporting related to the Alaska Alternate Assessment. In addressing technical documentation, we first present content evidence, then reliability, then descriptive statistics and AYP calculations.

In the end, both procedural and empirical evidence are brought to bear for supporting the claim that students with significant cognitive disabilities are assessed in a standardized system of reliable scoring and are achieving at various levels of proficiency on the alternate assessment.

Chapters 7 and 9 share the same set of appendices (Appendix 7.1 – 7.4). Chapter 8 includes strand, task, and item difficulty statistics within the body of the technical report.

In each appendix (7.1 through 7.4) statistics are presented in the following order:

- 7.1 Reading
- 7.2 Writing
- 7.3 Mathematics
- 7.4 Science

In the first four appendices in Chapter 7, descriptive statistics are presented in this order:

AYP

- Participation descriptive statistics at the total test level for each grade or grade band
- Score descriptives for total tests for each subject, including the frequencies of each score (some tables were overly lengthy and were thus not included).

Test Strand Descriptive Statistics

- Test Strand descriptives for each subject in the lowest grade or grade band (i.e., grade 3 for grade band 3/4).
- Strand descriptives for each subject in the next lowest grade or grade band (i.e., grade 5 for grade band 5/6).

Task Item Descriptive Statistics

- Operational task descriptives for each subject in the lowest grade or grade band (i.e., grade 3 for grade band 3/4).
- Task descriptives (includes operational and field test items) for each subject in the lowest grade or grade band (i.e., grade 3 for grade band 3/4).
- Task item descriptives for each subject in the next lowest grade or grade band (i.e., grade 5 for grade band 5/6).

Reliability

- Item reliability for each subject in each grade band (i.e., grade 3/4).

This pattern continues until the highest grade or grade band (10 or 9/10). The subjects always appear in the following order: reading, writing, math, and science.

The AYP tables depict percentages of students participating, then the number of students at each score value for the total sum in a subject area.

Strand, task, and item descriptive statistics tables depict the number of valid entries/items (N), the minimum and maximum values possible for items, the average score (Mean) and the average variation of scores around the mean (Standard Deviation).

The reliability section includes inter-item correlation matrices, when the size of the table allowed for inclusion in the appendix, average values (Mean), the range of scores (Variation), average variation around the mean (Standard Deviation), and the number of items measured (N). Cronbach's Alpha Based on Standardized Items statistics are also presented.

CHAPTER 1: BACKGROUND OF THE ALASKA ALTERNATE ASSESSMENT

Overview

The 2011–2012 Alaska Alternate Assessment represented an equivalent form test to the 2009–2010 Alternate Assessment. This version of the assessment is referred to as “Form A.”

History of Previous Program

In 2005, a Reliability and Validity study was conducted by Dr. Gerald Tindal which concluded that a need for revision to the State of Alaska’s Student Portfolio system in order to meet technical quality requirements set by the No Child Left Behind legislation was required. Following the department’s Request for Proposals, Dillard Research Associates was awarded a contract to secure a standardized performance-task assessment for students with significant cognitive disabilities. To provide greater reliability in administration and scoring of the assessment, an online administrator-training program was developed. This online training program includes training and proficiency tests for each subject area. Secure tests were developed in accordance with the State of Alaska’s Extended Grade Level Expectations (ExGLEs). Proficiency Level Descriptors (PLDs) were created by teams of content experts.

Current Program Overview

Reasons for Current Approach

The current Alaska Alternate Assessment system was developed to meet the requirements of the No Child Left Behind Act of 2001; as described by the National Center on Education Outcomes, alternate assessments are “tools used to evaluate the performance of students who are unable to participate in regular state assessments even with accommodations. Alternate assessments provide a mechanism for students with the most significant cognitive disabilities and for other students who may need alternate assessment formats to be included in the accountability system.”

The Alaska Alternate Assessments are standardized performance tasks administered and scored by Assessors who undergo a multi-step qualification process. Alaska’s current system of assessing students with significant cognitive disabilities has been approved by the U.S. ED Title 1 Final Assessment System Peer Review process.

Roles of Contractor, Department, and Others

The contractor, Dillard Research Associates (DRA), serves the Alaska Department of Education and Early Development (EED) in developing, training, administering, scoring, and data reporting related to the alternate assessment based on alternate achievement standards (AA-AAS) for students with significant cognitive disabilities. These tasks are defined in greater detail in subsequent sections of this Technical Report.

The EED maintains authority to finalize all deliverable documents, training systems, and reports stemming from the Alaska Alternate Assessment system. The contractor works closely and collegially with personnel in EED's assessment office.

Summary of Current Program

Description of Program

The Alaska Alternate Assessments are standardized performance tasks administered and scored by Assessors who undergo a multi-step qualification process. The Alternate Assessments are administered to students with significant cognitive disabilities in grades 3-10 (grade bands 3-4, 5-6, 7-8, and 9-10) and measure student achievement in relation to the ExGLEs. All students are assessed in reading, writing, and mathematics. Students in grades 4, 8, and 10 are also assessed in science. The Alaska Alternate Assessments focus on basic academic skills comprised of reading, writing, mathematics, and science tasks that are aligned with Alaska's ExGLEs.

The alternate assessments are comprised of the following components of a web-based training system that can be located at the following URL: <http://ak.k12test.com>

- Video-based training in each task
- Proficiency examinations
- Practice tests
- Secure test materials accessible only to qualified assessors during the test window
- A data entry and reporting portal

Description of Students Served

The Alaska Alternate Assessments are administered to students with significant cognitive disabilities in grades 3-10 and measure student achievement in relation to the ExGLEs. Each student's Individual Education Program (IEP) team determines which assessment students in Alaska's Statewide Assessment Program will participate in, based upon criteria established by the EED.

The Alaska Alternate Assessments focus on basic academic skills comprised of reading, writing, mathematics, and science tasks that relate to Alaska's ExGLEs. All eligible students are assessed in reading, writing, and mathematics. Students in grades 4, 8, and 10 are also assessed in science.

Description of How Scores Are Used

Assessors pre-enter their caseload of students into the online system. After administering the assessments one-on-one to a student, Assessors enter student scores directly into the online scoring and reporting system. An unofficial student report is immediately generated for the purpose of providing instructional feedback and guidance to IEP teams. Official student reports that have had the demographic information checked for accuracy and have been assigned proficiency levels were made available to districts on May 16, 2012 via the District Test Coordinators at the secure DRA Web Reporting System. These scores form the basis for Alaska's Adequate Yearly Progress (AYP) report for these students.

Any Standards-Based Assessment (SBA) and Alaska Alternate Assessment receiving a valid score in the content areas of reading, writing, math, and/or science will count toward overall participation and proficiency in the specific content area for AYP. Up to 1% of students attaining proficiency on the Alaska Alternate Assessment may count toward AYP proficiency per district.

Significant Changes Since Previous Technical Report

Improvements were made to several areas of the Alaska Alternate Assessment for the 2011-2012 testing window, including changes to training and the online training website. The improvements and changes are detailed in Appendix 1.1.

Procedural Changes

- Updated the Refresher Training Tasks, based on an analysis of Proficiency Test questions missed, and to reflect enhanced training in Writing Scoring and changes to the ELOS administration

Training and Continuous Improvement

- Developed comprehensive manual and workbook to improve the scoring of student writing samples
- Developed training supports for Mentors around the new Writing Scoring Manual
- Executed a Writing Scoring study

Specific ak.k12test.com Website Changes

- Organized the ELOS test documents into the matrix of Subject Area by Grade Band
- Developed enhanced visual representation on the training site of Full Training requirements and Refresher training requirements

Data Entry Tab

- Added a link to the Learner Characteristics Inventory (a data-collection tool for the GSEG) on the Student Entry page

Admin Tab

- Added a fifth column to the display of data on the Data Entry Status report. The fifth column indicates how many of an Assessor's student scores have been submitted to EED

*Appendix 1.1 Web Changes Handout***Extended Levels of Support (ELOS) test documents**

- Developed all new ELOS test items
- Aligned to Early Entry Points to the Extended Grade Level Standards
- Grade-banded all subject areas (3/4, 5/6, 7/8, and 9/10)
- Reduced all tests to three tasks of five items each
- Ordered the five items within each task to: Attention item, Interaction item, Easy item, Medium item, and Hard item.
- Revised administration rules to require administration of all 15 items

*Appendix 1.2 ELOS Item Maps 2012***Organization of Technical Report**

The 2012 Technical Report is organized around ten broad topics, with detailed appendices referenced where appropriate. The Report serves as a narrative description of the activities and results of the 2011-2012 testing year. The appendices provide all reference materials, including training agendas, guidance documents, and complete statistical analyses on a variety of required reporting topics.

The topics of the Technical Report are:

Background of the Alaska Alternate Assessment
Test Design and Item/Task Development
Test Administration Procedures
Scoring
Standards Validation
Reporting
Test Validity
Descriptive Statistics
Adequate Yearly Progress (AYP)
Recommended Program Improvements

In all sections where subject area results are reported or described, the Technical Report standardizes the reporting order to: Reading, Writing, Mathematics, and Science.

CHAPTER 2: TEST DESIGN AND ITEM/TASK DEVELOPMENT

Overview

The National Center on Education Outcomes describes alternate assessments as "tools used to evaluate the performance of students who are unable to participate in regular state assessments even with accommodations. Alternate assessments provide a mechanism for students with the most significant cognitive disabilities and for other students who may need alternate assessment formats to be included in the accountability system."

<http://www.cehd.umn.edu/NCEO/TopicAreas/AlternateAssessments/altAssessTopic.htm>

The need for developing alternate assessments was in line with the requirements of the Goals 2000 and Improving America's Schools Act (IASA), the Individuals with Disabilities Education Act of 1997 (IDEA), and the IDEA reauthorization in 2004, as well as Alaska's Quality Schools Initiative (QSI), which supported high standards, statewide assessments, and improved results for all students. Until mandated by the federal government, most students with significant cognitive disabilities, and other students with disabilities, were excluded from assessment. Alternate assessments are not typical large-scale assessments, nor are they individualized diagnostic tools. However, the goal is to provide information and accountability for the academic performance of all students in a school district.

The original design of Alaska's Alternate Assessment, a student portfolio, was intended to provide an accountability measure that was consistent with state standards, individualized, performance-based, used independent and reliable scoring, and could be integrated with curriculum and the student's IEP. Students were assessed in language arts, mathematics, and skills for a healthy life. The portfolio assessment was very time consuming for teachers, and teachers often felt that the portfolio measured their ability to construct a portfolio rather than what a student was learning. However, many of the purposes of this first alternate assessment were met. Students were included in the state's comprehensive system of student assessment; student IEPs used academic content standards as goals; students were assessed on academic progress; and, students were included in general education classrooms on a more frequent basis.

After conducting a reliability and validity study, Alaska moved to a performance task assessment that focuses on measuring reading, writing, mathematics, and science. The current Alaska Alternate Assessment uses performance tasks to measure what a student knows and can do in those four core subject areas. The state felt that an assessment with performance tasks offered a more standardized assessment with high technical quality (reliability and validity). Generally, surveys of teachers indicate a greater overall satisfaction with the performance task assessment.

Description of ExGLEs and their relationship to GLEs

In 1993, the EED developed content standards in English, mathematics, science, geography, history, skills for a healthy life, government and citizenship, fine arts, technology, and world languages. The content standards were broad statements of what students should know and be able to do as a result of their public school experience. A revised edition included content standards for employability, library information/literacy, and cultural standards for students. These content standards are discussed in this document as Grade Level Expectations (GLEs).

In 1999, the Alaska State Board of Education adopted extended performance standards for students with significant cognitive disabilities in the content areas of English/language arts, math, and skills for a healthy life. The reason for developing extended performance standards was to allow for variation in the demonstration of skills across ages and abilities. Different content standards were assigned to, and assessed at, different grade levels.

In response to the 2001 No Child Left Behind legislation, a third edition of the Alaska content standards booklet includes expanded performance standards organized by grade band, called ExGLEs, and revised science content standards and science extended performance standards by grade band. A fourth publication included Alaska history standards.

The No Child Left Behind legislation also required that if a state used AA-AAS's for students with significant cognitive disabilities, "the assessment materials should show a clear link to the content standards for the grade in which the student is enrolled although the grade-level content may be reduced in complexity or modified to reflect pre-requisite skills." In response to this section, the Alaska EED began the process of developing ExGLEs and PLDs.

The ExGLEs are an interpretation of the content standards that should be taught and learned within each grade level. The content is reduced in complexity to provide entry points to the GLEs, while still providing challenging academic expectations for students with significant cognitive disabilities.

Test Specifications and Blueprint

Description of Test Specifications (DOTS)

Descriptions of Test Specifications for the 2011–2012 Alternate Assessment are Excel spreadsheets that define all aspects of each item used in the all test materials. In addition to items used in the 2011–2012 assessments, information related to all items used in tests beginning with the 2007–2008 test materials are displayed. Information includes the strand name, the number of answer options, maximum score points, item depth of knowledge (DOK), whether the item was an operational or field test item, and statistical data for each item (mean, standard deviation), a statistical analysis of the difficulty of the item (the mean points for each item divided by the maximum points available), and the task weight.

The DOTS documents for reading, writing, mathematics and science contain confidential secure test information and are not available to the public.

Process of Establishing Test Specifications

The test specifications included the following variables as items were developed:

Grade Level – All items were written to appropriate grade bands: 3-4, 5-6, 7-8, 9-10.

Subject – All items were written within specific subject area domains: Reading, Writing, Mathematics, and Science.

Strand Name: All items were written to fit within subject domains.

Extended Grade Level Expectation: These expectations within a content area were organized in content strands and used to organize item writing.

Item Prompt: Each item included specific wording for the teacher to use in test administration.

Item Type: Both selected and constructed-response items were considered with the vast majority of items using selection responses so that students with physical limitations could participate (respond).

Item Answer: Each item was constructed with three options if using a selection type response or an area for the student to construct a response.

Bias / Content Panel Judgment: Committee members rated each cousin item as Easy (E), Medium (M) or Hard (H) for students taking the Alaska Alternate Assessment.

Item Depth of Knowledge:

- Level 1 Rote memory, recall, simple procedure, or apply a one-step, well-defined algorithmic procedure (identify, recall, recognize, use, measure).
- Level 2 Some mental processing beyond habitual response. Decisions in how to approach a problem (classify, organize, estimate, display data, compare data).
- Level 3 Reasoning, planning, using evidence -- complex and abstract (draw conclusions, cite evidence, explain in terms of concepts, decide which concepts to apply to solve a complex problem). More than one answer, and student has to justify their response.
- Level 4 Complex reasoning, planning, developing and thinking, most likely over an extended period of time, plus applying significant conceptual understanding and higher-order thinking. Make several connections (relate ideas within the content

area or among content areas, and select one approach among many alternatives to solve the problem). Design and conduct experiments and projects, develop and prove conjectures, make connections, combine and synthesize ideas into new concepts, critique experimental designs.

Item Content Test Blueprint and Item Specifications

Test construction for the 2012 testing window was designed to closely match the Form A test administered in 2010. The percent of the strands represented in each subject area and at each grade band are displayed in the appendix.

Appendix 2.1 Strands: Reading, Writing, Mathematics, and Science

Linear Equating

The Reading, Writing, and Mathematics Alternate Assessments tests had significant changes from 05-06 to this year, and the scores are calculated between the two to reach established standard setting cut scores. The point totals of the 2005-2006 assessments available within each strand were established as the original year. Point totals within each strand in the 2009-10 assessments were compared to the original year, and a weighting factor calculated.

For instance, assume the 2006 strand 1.34 Numeration totaled 30 possible points and the point total in the same strand for the 2012 test totaled 28, the weighting factor would be 1.07.

Because the science test was first employed in 2008-2009 and is not based on the 2005-2006 assessment (and the number of points available are the same from 08-09 to 11-12), the Science Alternate Assessment didn't need to undergo linear equating.

Appendix 2.2 Weights: Reading, Writing, and Mathematics

Proficiency Level Descriptor Development

Prior to the adoption of the new Alaska Alternate Assessment, the extended performance standards needed to be revised to reflect the change in the general education academic standards. The existing proficiency level descriptors for the Alternate Assessment Portfolio were universal descriptors. The department assembled teams of content and special education experts, as well as other stakeholders, for the purpose of developing Extended Grade Level Expectations (ExGLEs) for the grade bands 3-4, 5-6, 7-8 and 9-10, and grade-banded Proficiency Level Descriptors based on alternate achievement standards (PLDs) for students with significant cognitive disabilities. The Official Individual Student Reports (ISRs) contain the definitions and descriptions for each proficiency level and at each grade level for each subject area Alaska Alternate Assessment.

Appendix 2.3 Proficiency Level Descriptors

Cut Scores

A standard-setting committee determined cut scores for the new alternate assessment and used the PLDs during that process. During standard setting, the PLDs were revised and were formally adopted by the State Board of Education in July 2007 (reading, writing, and mathematics) and in July 2008 (science). To obtain a proficiency level of advanced, proficient, below proficient, or far below proficient in reading, writing, and mathematics on the Alaska Alternate Assessment, a student must obtain a score as set out in the following tables:

| Reading Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|----------------------------------|------------------------|------------------------|------------------------|-------------------------|
| Advanced | 63 or above | 77 or above | 52 or above | 57 or above |
| Proficient | 32-62 | 46-76 | 33-51 | 43-56 |
| Below Proficient | 8-31 | 11-45 | 12-32 | 22-42 |
| Far Below Proficient | 7 or below | 10 or below | 11 or below | 21 or below |

| Writing Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|----------------------------------|------------------------|------------------------|------------------------|-------------------------|
| Advanced | 76 or above | 67 or above | 76 or above | 82 or above |
| Proficient | 38-75 | 33-66 | 41-75 | 47-81 |
| Below Proficient | 7-37 | 10-32 | 16-40 | 24-46 |
| Far Below Proficient | 6 or below | 9 or below | 15 or below | 23 or below |

| Mathematics Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|--------------------------------------|------------------------|------------------------|------------------------|-------------------------|
| Advanced | 62 or above | 61 or above | 74 or above | 81 or above |
| Proficient | 33-61 | 25-60 | 52-73 | 63-80 |
| Below Proficient | 6-32 | 8-24 | 22-51 | 24-62 |
| Far Below Proficient | 5 or below | 7 or below | 21 or below | 23 or below |

| Science Proficiency Level | Grade 4 | Grade 8 | Grade 10 |
|----------------------------------|----------------|----------------|-----------------|
| Advanced | 44 or above | 44 or above | 44 or above |
| Proficient | 24 - 43 | 29 - 43 | 26 - 43 |
| Below Proficient | 12 - 23 | 16 - 28 | 18 - 25 |
| Far Below Proficient | 11 or below | 15 or below | 17 or below |

In addition to the cut scores established above, EED also determines AYP for English language arts (ELA) using a combined Reading and Writing score. The lowest possible proficient scores in each category are added together to form the ELA cut score. There are two performance levels, Above or Below, as shown below:

| ELA Proficiency Level | Grade 3 & 4 | Grade 5 & 6 | Grade 7 & 8 | Grade 9 & 10 |
|------------------------------|------------------------|------------------------|------------------------|-------------------------|
| Above | 70 or above | 79 or above | 74 or above | 90 or above |
| Below | 69 or below | 78 or below | 73 or below | 89 or below |

Item/Task Development

Item Writing, including Scoring Guides

A robust set of field test items were designed in 2009-2010 and underwent Content and Bias Review; no new items were written for the 2012 test window. The 2011-2012 Alaska Alternate Assessment test documents were identical to the test documents deployed in 2009-2010, and are referred to as "Form A."

The training for scoring writing samples was enhanced for the 2012 test window. New guidance tools included a Writing Scoring Manual, Writing Scoring Mini-Lesson and Answers, Writing Scoring Activities, and enhanced online training.

Appendix 2.4 Writing Training Documents

Item Development and Expanded Levels of Support (ELOS)

Between June and September 2011, the ELOS test documents were substantially improved. Previously, one set of documents covered the Extended Levels of Support administration for all four grade bands in reading, writing, mathematics and science, respectively. The current deployment of the ELOS administration includes new test documents for each subject area at each grade band level.

Previous administration rules required Assessors to locate appropriate tasks from a wide range within each test document and to administer a minimum of fifteen items. Administration codes included the opportunity to record items the student already mastered (A), inappropriate (I), or refused (R), as well as indicating the level of support required to bring the student to success on each item (1 = full physical support, 2 = partial physical support, 3 = visual, verbal, and/or gestural prompts, 4 = student completes task independently).

The new ELOS administration documents are limited to only three tasks of five items each at each grade level and in each subject area. Within each task, the five items are ordered to provide an attention item, an interaction item, an easy item, a medium item and a hard item. Administration rules were revised to require Assessors to administer all fifteen items in the test document for those students who qualified for the ELOS administration in a particular subject area.

New ELOS Test Booklets

Based upon longitudinal input gathered from practitioners and stakeholders, EED requested the development of new ELOS assessments. The ELOS development plan to create new Scoring Protocol booklets (SPs) and Student Materials booklets (SMs) went through several iterations in order to provide a plan that made the items and tasks most accessible to students with the most significant access limitations while ensuring alignment to the Early Entry Points (EEPs) and Extended Grade Level Equivalents (ExGLEs), where

relevant. All graphics were reviewed and improved in terms of age-appropriateness, clarity, and accessibility.

Test structure was also redefined for all new tasks and items. The final plan, which was adopted by EED in June of 2011, is elaborated below. Examples of items that convey how alignment and cognitive complexity are addressed are provided thereafter.

The rating scale that was in used in prior years was modified due to the increased access provided by the new task design. As mentioned above, the scale for "A" – Student already has this skill, "I" - Inappropriate/Inaccessible based on the nature of the student's disability, and "R" – student refuses to complete are no longer scoring options for ELOS tasks and items. Students are scored in accord with the Levels of Independence Scoring Rubric:

1. Full Physical Contact needed for response
2. Partial Physical Contact needed for response
3. Visual, Verbal, and/or Gestural support needed for response
4. Independent, no contact or prompting needed for response

Final Plan

DRA employed the following plan in the development of new ELOS test materials (Student Materials & Scoring Protocols) in all four subjects (reading, writing, math, and science) across all four grade bands (3/4, 5/6, 7/8, & 9/10; *science at grades 4, 8, & 10*):

1. Task and Item Content

Task and item content were designed to align to the Early Entry Points (EEPs) in the early grade bands (3/4, 5/6, and 7/8). Some difficult items at the upper grade band (9/10) align to an Extended Grade Level Expectation (ExGLE). In order to increase accessibility and familiarity, tasks and items also include instructional materials that the student should be familiar with in order to decrease the novelty of the testing situation. For example, students are asked questions about their friends, their favorite picture book, their address, their school, the weather that day – topics that are relevant and immediate to their specific surroundings.

2. Test structure

ELOS tests were composed of three tasks, with five items per task, in each content area (reading, writing, math, science). This results in the administration of 15 items per content area. Assessors administer all tasks, all items on the ELOS administration for students who are eligible.

| Grade/ Subject | Reading | Writing | Math | Science | Total |
|-------------------|---------|---------|------|---------|-------|
| 3 | 15 | 15 | 15 | 0 | 45 |
| 4 | 15 | 15 | 15 | 15 | 60 |
| 5 | 15 | 15 | 15 | 0 | 45 |
| 6 | 15 | 15 | 15 | 0 | 60 |
| 7 | 15 | 15 | 15 | 0 | 45 |
| 8 | 15 | 15 | 15 | 15 | 60 |
| 9 | 15 | 15 | 15 | 0 | 45 |
| 10 | 15 | 15 | 15 | 15 | 60 |

3. Task and Item Structure

DRA developed tasks that are arranged in the following structure: the first item within each task will be an *Attention* item. For example,

"Here is a book." (student attends to book presented by Assessor)

The second will be an *Interaction* item.

"Touch this book." (student interacts with book)

The 3rd, 4th, and 5th were organized by content and cognitive complexity in the categories of *Easy*, *Medium*, and *Hard*, respectively. It is at times more important to have sufficient coverage of an EEP than it is to have items that are clearly articulated along this continuum of complexity within a specific task. Tasks are also vertically articulated across grade bands with subtle changes in task difficulty from grade band to grade band.

As an example of how items are developed along this continuum, consider the following example. An easy level item for a 3/4 grade band spelling task may begin with asking the student to copy the word "cat", with a model provided. This is common term in most classrooms and is composed of a small number of letters, only three. In addition, it is a high interest concept for most students in this domain, who tend to be fascinated by animals. A medium item in the same task might ask the student to spell the word "dog" with no model provided. Again, utilizing a word that the student has likely seen or heard multiple times throughout his/her schooling experience, but a word that is simple and composed of only three letters. The student must generate the word without a model, which makes it more difficult than the easy example. A hard item in this task might be to ask the student to spell the word "bear". It is easy to see that complexity increases on an item-by-item level based upon supports provided, content, and measureable aspects such as letter count.

4. Alignment to EEPs, ExGLEs, and Strands

Themes were used to organize each task (e.g., the items in one task will have one general concept, such as 'polar bear,' and the five items will target different strands within one task). As mentioned, items in the lower grade bands (3/4, 5/6, and 7/8) are aligned to the EEPs. Some difficult items at the 9/10 level are aligned to lower level ExGLEs. The new item maps identify the content standard, strand, ExGLE, and EEP of each item independently.

5. New ELOS items were developed using DRA's internal item development process, with EED review and certification.

Employ DRA's internal item-development process. Item development was completed at DRA in June and July of 2011. All tasks and items were reviewed by independent proofreaders, as well as DRA and EED staff on multiple occasions.

6. New ELOS test books (Scoring Protocols and Student Materials) in each of four subject areas (reading, writing, mathematics, and science) and across four grade bands (3/4, 5/6, 7/8 & 9/10; *science at grades 4, 8, & 10*) were developed and finalized in January 2012.

Task/item prompts are included in the Student Materials booklets. Scoring Protocols include the Preparation Guides and the task/item prompts, student response expectations, and scoring levels.

7. ELOS Item Maps were updated to match the final ELOS task/item development. New ELOS items were transferred to the Directory of Test Specifications (DOTS) in February 2012.

8. The new ELOS test booklets were implemented in 2012 as a field test. The purpose of administering the ELOS items is to allow access to the assessments and support participation in statewide testing events. ELOS results are also used for instructional purposes; however, they are not used for AYP reporting purposes. The student's data from the administration of the Alaska Alternate Assessment standard items are used for reporting both performance and participation as part of AYP. As such, implementing a live field test is appropriate. Input regarding the assessments was solicited and gathered from Assessors regarding the changes that were made to the test structure and the impact upon the classroom as part of the *Consequential Survey*. Input will also be requested during trainings.

9. The *Materials Preparation Guide* was developed, inclusive of all preparation guides used for each test item across both the Standard and ELOS administrations. This guidance document was provided to Mentors over two weeks in advance of the spring test window and distributed to Assessors so they can prepare all of the materials they need to gather to administer the alternate assessments, especially the newly designed ELOS tasks/items.

Scoring Protocol and Student Material Examples

The following is an example of a task written in the domain of mathematics.
ELOS Math Scoring Protocol Example Task:

| Math Grades 5 & 6 - Task 1 | | | | |
|--|---------------|---|---|---|
| ELOS Skills | Student Score | | | |
| 1. <i>Attention</i> : " Here are several coins. " (student attends to coins presented) | 1 | 2 | 3 | 4 |
| 2. <i>Interaction</i> : " Touch this nickel. " (student safely interacts with nickel) | 1 | 2 | 3 | 4 |
| 3. <i>Easy</i> : " Which of these coins is a quarter? " (student identifies quarter) | 1 | 2 | 3 | 4 |
| 4. <i>Medium</i> : " Which group has more, the dimes or pennies? " (student identifies dimes) | 1 | 2 | 3 | 4 |
| 5. <i>Hard</i> : " Which coin is worth more, the dime or the penny? " (student identifies dime) | 1 | 2 | 3 | 4 |

The attention items were designed such that the student's attention is directed to a content-related object, in this example several real coins. The interaction item provides the student with a tangible direction, which the student attempts to follow. Item 3 provides a content prompt asking for a simple coin identification when provided a real object. Item 4 asks the student to compare a group of three dimes and two pennies to determine which group has more coins. This is a multi-step operation for the student (counting two sets and making a comparison), making Item 4 more difficult than Item 3. Item 5, the most difficult item on this task, asks the student to compare how much a dime is worth versus a penny. Adding to the difficulty of item 5 is the fact that a dime is actually physically smaller than a penny though it represents a larger number of cents. Students in this domain are not adept abstract thinkers, which makes this item clearly more difficult than the others.
ELOS Math Student Materials Example Task:

| Math Grades 5 & 6 - Task 1 |
|----------------------------|
|----------------------------|

1. Assessor places a quarter, dime, nickel, and penny within the student's field of vision/access.
2. Assessor provides student with the nickel (Assessor will need to ensure that interaction is safe for students who put objects in their mouths).
3. Assessor places quarter and dime within the student's field of vision/access.
4. Assessor places three dimes and two pennies within the student's field of vision/access.
5. Assessor places a dime and a penny within the student's field of vision/access.

The materials provided for this task are real objects (quarter, dimes, nickel, pennies), which is consistent with the new test design. Having the students work with real objects makes this task more accessible, more relevant, and less novel.

Reduction in Complexity, Depth, and Breadth

Due to the federal regulations provided in December 2003, steps were taken to increase the cognitive accessibility of items. This was done by analyzing and removing potential barriers for students with significant cognitive disabilities. This process was used in the development of items and for both administration and scoring and student materials. Simplified language was used in all text. Alignment was ensured between teacher-scripted language and student materials. General test layout was considered from the view of readability and legibility. Specific administration directions were limited to a single page of the Scoring Protocol for ease of administration. Pictures were constructed using primarily black and white for minimal complexity. All items were reviewed with administration and development steps toward reducing complexity.

Reductions in depth, which is generally defined by Anderson's revision of *Bloom's Taxonomy*, were accomplished by limiting the process verbs to simpler tasks (recognize, identify, match, understand *versus* analyze, develop, evaluate, create). The team developed items that linked to the relevant ExGLEs in reading, writing, mathematics, and science at the grades tested. From that point, the teams tried to target performance events that were reduced in terms of depth, but maintained access to appropriate content.

Reductions in breadth, which can be defined in terms of how broad a student's domain of knowledge must be to answer a specific item, were accomplished by limiting the item content to accessible domains. For example, while a general education assessment might target the process of implementing a laboratory experiment in science, the alternate assessment might ask the student to define a term that is critical to the experiment. The content is relevant, but the performance demand does not require a wide knowledge set to answer appropriately.

Reductions in complexity, which is generally how difficult the test content is, were accomplished by limiting the difficulty of the content (e.g., adding single-digit integers is much easier than adding imaginary numbers, though the process verb, "to add", is the same). Language load was also analyzed and decreased in order to increase accessibility using the *Linguistic Complexity Rubric for Universal Design* (Instrument 1).

It is critical to mention that depth, breadth, and complexity are intertwined and work together to determine overall item difficulty. They are simply three lenses used to systematically address and make items more accessible from a test content perspective.

As mentioned, tasks and items were developed based on a one-to-one correspondence with the ExGLEs. All strands and attributes were equally addressed in accordance to proportion of points for each task. The total points for each test was fixed at 100 points to allow proficiency standards from the first year to be comparable to the second year of testing. Weighting was needed and an algorithm was used to equalize the differential points across strands/attributes.

Depth-of-knowledge (DOK) was judged in the analysis of the Alaska Alternate Assessment. An alignment study was conducted in 2007 by Karvenon and Almond; the information was used to guide item adaptations for the 2007-2008 secure test items. Categorical concurrence, range of knowledge, and balance of representation were defined originally by Webb, and adapted by Dr. Tindal for use with students with significant cognitive disabilities, and then defined based on operational use within the Alaska Alignment Study.

New ELOS items were developed this year using a similar approach, yet tasks and items were developed based on a one-to-one correspondence primarily with the Early Entry Points (EEPs), which are the prerequisite skills a student needs to access the ExGLEs. At the high school level, some ELOS items were developed with a one-to-one correspondence with the ExGLEs in order to provide for an increased range of difficulty as students progress through the grade bands. Field test results will be analyzed to determine item characteristics. Bias and sensitivity concerns will be addressed via the *Consequential Survey* results related to the new ELOS design and items.

Bias and Sensitivity Review

A bias and sensitivity review of the new test items was conducted in November 2007. During this process, reviewers examined the bias of the assessment and if the format would affect student performance. A group of 12 participants from Alaska and two specialists with the deaf and blind community from Oregon were selected to review all items. All reviewers were given examples to focus on during the review and all held Qualified Assessor certificates and certification in special education. Items were updated based on the results of this review prior to the 2007-2008 testing window. Bias and sensitivity were analyzed for all items using the *Bias and Sensitivity Review Checklist* (Instrument 2). Items were either adapted to meet the groups concerns or not utilized on the assessment based upon the results garnered.

A second Bias and Content Committee was convened in September 2009 to analyze cousin items, a pool of new, related items to the existing items in the Alaska Alternate Assessment. The results of this analysis are indicated in the DOTS document.

Appendix 2.5 Instrument 1 - Linguistic Complexity & Instrument 2 – Bias and Sensitivity Review

Test Design and Development

Representation and Functionality

The 2009-2010 cousin items reflected minimal construct under-representation or construct irrelevant variance (CIV) to ensure functionality.

- Select the most appropriate word with the least number of syllables
- Reduce number of words used in items, directions, and passages
- Use independent clause structure instead of dependent clause structure in passages
- Develop prompts with minimal wording
- Ensure more opportunities for modeling
- Provide examples when possible
- Create clear (not tricky) distractors
- Provide explicit textual information with reduced requirements for extended inference
- Provide rules rather than exceptions
- Use careful sequencing so that potentially similar/confusing information is not presented
- Place items adjacent to similar information
- Provide multiple choice options for items when possible or appropriate for item construction

Rasch equating was used to ensure functionality by calculating fit statistics that reflect the degree to which ability and difficulty are mapping correctly.

Psychometric Guidelines for Selecting Items/Tasks for Item/Task Bank

Traditional guidelines were used for selecting items and tasks that rely on reliability coefficients but also on implementation in the field. As described in other sections of this report, extensive training of new and returning Qualified Assessors and Mentors was conducted before the testing window opened; in addition, web-based training and proficiency assessments were completed with actual practice in the field required.

We calculated both the mean and standard deviation for each item to ensure the item was functional for a wide range of students.

The entire item bank was developed with all items from the original test completed in 2006-2007 and every year after that to identify common items and use them as anchors for calibrating item values using a Rasch Partial Credit Model.

Item Bank Summary

All items that have been used in any version of the Alaska Alternate Assessment, beginning with SY 2006, have been included in the 2012 DOTS, due to be released to EED in November 2012. Student performance on each item is recorded for each year the item was in use. The DOTS is the item bank.

Current Test Construction

The 2008-2009 Alaska Alternate Assessment served as the baseline document for developing two forms; items included in this version are referred to as operational items. Test items in Form A (2009-2010) were developed by identifying the strand, task, and construct for each operational item and locating matching cousin items for each. This system allowed DRA to conduct statistical analysis on the operational items, on the field test items, and equivalent test form analysis. Beginning with testing year 2011-2012, all items in the current item bank are considered operational items.

Construction of the Operational Forms

Approximately one-half of the 2008-2009 test items (operational items) were replaced by matching cousin items (field test items). Operational items are items that have been used (without modification) in the 2007-2008 and 2008-2009 assessments, and thus have two years of statistical data collected for each item. Field test items were carefully created to match the operational items they would replace.

All test items for 2011-2012 (Form A) are operational items.

In 2009, DRA and EED constructed a plan that would allow a minimum of six versions of the Alaska Alternate Assessment:

- 1) AK AA Test (2007-2008, 2008-2009)
- 2) 1/2 (a) AK AA test + 1/2 (a) FT (FORM A, 2009-2010)
- 3) 1/2 (b) AK AA test + 1/2 (b) FT (FORM B, 2010-2011)
- 4) 1/2 (a) AK AA test + 1/2 (b) FT
- 5) 1/2 (b) AK AA test + 1/2 (a) FT
- 6) 1/2 (a) FT + 1/2 (b) FT

However, because EED joined a General Supervision Enhancement Grant (GSEG) consortium of states working to create alternate assessments, DRA and EED have agreed to maintain the current two forms of the Alternate Assessment (Form A and Form B, numbers 2 and 3 above) for the future, until such time as the General Supervision Enhancement Grant (GSEG) assessment is developed and ready to be deployed.

- 1) Form A, 2011-2012 (described in #2 above)
- 2) Form B, 2012-2013 (described in #3 above)

Test Development Timeline

At the conclusion of the first six-year contract with EED on June 30, 2011, DRA had developed and produced two complete forms of the Alaska Alternate Assessment. Both forms have approximately the same number of tasks and items and represent similar content standards and strands. In addition, a new ELOS test was developed, and implemented in the 2012 test year.

The new six-year contract, representing 2011-2017, was established as a maintenance contract, with no new test items to be developed.

Possible Future Test Construction

Psychometric Guidelines for Constructing Future Forms

In 2010-2011, DRA conducted an analysis of item functioning in anticipation of reconstructing the Alaska Alternate Assessments into equally weighted tests, in order to eliminate the need for linear equating of scores. However, in light of Alaska's involvement in a General Supervision Enhancement Grant (GSEG) consortium, both EED and DRA agreed to continue use of the current Form A and Form B test documents in anticipation of a new test design produced by the GSEG, as described in the previous pages. The following information is included as a reminder of the analyses that were conducted.

Possible Psychometric Guidelines for Constructing Future Forms

Coverage of Strands (and equal weighting) was used to ensure appropriate and consistent representation of items to strands within each subject area and grade level.

Two forms were developed with 2009-2010 – Form A (where we used half of the items from previous years to serve as anchored operational items and developed cousin items for the other half of the test) and the 2010-2011 – Form B (where we again used half of the items from previous years to serve as anchored operational items and developed cousin items for the other half of the test). In all tests, the cousin items were embedded in the actual test but only the operational items were used to calculate Adequate Yearly Progress (AYP).

During the summer of 2011, all items were calibrated to a common scale across years using a Rasch Partial Credit Model (RPCM). The results of the scaling analysis provided information on all item difficulties and functioning. This information can be used to construct alternate forms of equivalent difficulty. During the scaling analysis, all tests were equated between successive years with a nonequivalent group with anchor test (NEAT) design. Our intent was to produce item calibrations for the 2009-2010 (Form A) and 2010-2011 (Form B) forms that were equated to the original year of each subject's testing– the first year the test was administered. We used a chained equating method to link items between multiple years. Our method comprised the following steps:

- Calibrate all items in the original year with a RPCM model
- Identify common items between original and subsequent years
- Calibrate subsequent years while anchoring common item difficulties and step values to the original calibrations

When equating item calibrations in 2009-2010 and 2010-2011 to the original year scale, we used any and all common items across years as anchor items, anchoring common items

between adjacent years and from previous years. This process allowed more items to be anchored, which ultimately should reduce the equating error.

Table 1 displays the original year for each subject and grade. Following Table 1 is a description of the common item-anchoring plan for the two different original years.

Table 1 Original Equating Year by Subject and Grade

| Original | Reading | Writing | Math | Science |
|------------|-----------|-----------|-----------|-----------|
| Grade 3-4 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |
| Grade 5-6 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |
| Grade 7-8 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |
| Grade 9-10 | 2006-2007 | 2006-2007 | 2006-2007 | 2007-2009 |

Original Test Year 2006-2007. Items were first calibrated with a RPCM during 2006-2007, the original year for determining linear equating in future years. After calibration, common items between 2006-2007 and 2007-2009 were identified (the same test had been used for two successive years). The 2007-2009 items were then calibrated with the common item difficulty and step values anchored to the 2006-2007 calibrations, which adjusted the calibration of the freely estimated 2007-2009 items *relative to* the 2006-2007 values. In 2009-2010 common items were identified between 2006-2007 and 2009-2010, and 2007-2009 and 2009-2010. The 2009-2010 items were then calibrated with the common item difficulty and step values anchored to the values from the year in which they originally appeared (either 2006-2007 or 2007-2009). In 2010-2011 common items were identified between 2006-2007, 2007-2009 and 2010-2011, and 2009-2010 and 2010-2011. The 2010-2011 items were then calibrated with the common item difficulty and step values anchored to the values from the year in which they originally appeared (either 2006-2007, 2007-2009, or 2009-2010).

Original Test Years 2007-2009. Because science was not a required for the Alaska assessment prior to 2007, the original years for these assessments were 2007-2009. After the initial calibration for the 2007-2009 years, common items were identified between 2007-2009 and 2009-2010. The 2009-2010 items were then calibrated with the common item difficulty and step values anchored to the 2007-2009 calibrations. In 2010-2011 common items were identified between 2007-2009 and 2009-2010, and 2010-2011. The 2010-2011 items were then calibrated with the common item difficulty and step values anchored to the values from the year in which they originally appeared (either 2007-2009 or 2009-2010).

Essentially equivalent items. When the original year was 2006-2007, there were occasionally no common items between the 2006-2007 version of the test and the 2007-2008 version of the test. Yet, our intent was to equate items to the original year, 2006-2007. If no common items existed between 2006-2007 and 2007-2008, we identified *essentially equivalent* items. For example, in the following two tasks, the items from 2006-2007 and 2010-2011 consisted of copying letter names; we simply associated items with

each other without exact matching of each specific letter with itself across the two time periods.

| 2006-2007 | | | | 2010-2011 | | | |
|-----------|--------|------------------|--------|--------------------------------|--------|------------------|--------|
| Item | Letter | Student Response | Points | 1.34A - Copy Letters - Scoring | | | |
| | | | | Item | Letter | Student Response | Points |
| 1 | g | | /2 | 1 | c | | /2 |
| 2 | t | | /2 | 2 | f | | /2 |
| 3 | k | | /2 | 3 | h | | /2 |
| 4 | a | | /2 | 4 | z | | /2 |
| 5 | m | | /2 | 5 | w | | /2 |
| 6 | B | | /2 | 6 | L | | /2 |
| 7 | C | | /2 | 7 | U | | /2 |
| 8 | W | | /2 | 8 | J | | /2 |
| 9 | I | | /2 | 9 | Y | | /2 |
| 10 | Z | | /2 | 10 | V | | /2 |
| | | | | 11 | G | | /2 |

Within Appendix 2.6, Tables 1-6 present the “equating map” used for reading by grade-band, Tables 7-10 present the same equating map for writing, Tables 11-14 present the map for math, and Tables 15-16 present the map for science. When viewing the tables, each row represents a unique item. When the item was presented in multiple years, its variable label appears in each year. The column containing the words “recode” indicates the way the scoring was recoded and was left blank if no recoding was deemed necessary. *Essentially equivalent* items are displayed in bold-faced font.

Appendix 2.6 Equating Map

Data Preparation

Five years of Alaska Alternate Assessment data were prepared for the equating analysis (using the RPCM methodology described above) with data prepared in a similar format for all grade-bands and for all content areas. A number of steps were consistently applied across the subject areas and grade levels that involved cleaning up the data file to exclude missing fields (values) for virtually all items for a small group of students, data aggregation to prepare master files for analysis, and item recoding for ensuring similar scales.

Systematic deletion of non-responders

The Alaska Alternate Assessments contain discontinuation rules for students who are not responding to items aligned with the ExGLEs. The discontinuation rules state that if a student scores a zero on three items within a task, the remainder of the task should be discontinued and the tester should go on to the next task. If the discontinuation rules are exercised for three consecutive tasks, the test as a whole should be discontinued. The student subsequently is administered the appropriate set of ELOS tasks/items. These rules were in place in the beginning of the 2007-2008 school year.

Data Aggregation

The alternate assessments were identical for all content areas (reading, writing, math, and science) during the 2007-2008 and 2008-2009 school years. All items were presented in the same format, in the same order, and with the same administration procedures. When applying item response models, such as the Rasch model, larger sample sizes are desirable, as the error associated with each item and person estimate included in the analysis are

reduced. Given the identical design of the assessments between 2007-2008 and 2008-2009, and the need for large sample sizes, all item response data were combined between these two years, creating a single “2007-2009” data file.

Item Score Recoding

There were four reasons that items occasionally needed to be recoded when the item included: (a) a large number of potential scoring options, (b) inconsistent scoring between years for common items, (c) inconsistent item scoring, or (d) missing step values.

In some content areas, particularly writing, a large amount of score reporting options were available. For example, one writing item in grade-band 9/10 had possible scores ranging from 0-35. When such a large number of scores were possible, it became difficult to estimate the step calibrations because very few students received each possible score. When a large number of scores were possible, the item was recoded into a narrower band of values. For example, the aforementioned 35-point item was recoded to a 7-point scale, with the score options categorized into five 6-point interval ranges and one 5-point interval range (i.e., 0 = 0, 1-6 = 1, 7-12 = 2, 13-18 = 3, 19-24 = 4, 25-30 = 5, and 31-35 = 6).

In other instances, the item scoring changed between years, although the item itself did not change. For example, an item may have been scored dichotomously, 0-1, in one year but then given a partial credit rating, 0-1-2, in the next. If the scoring changed on a common item that needed to be anchored, the items needed to be recoded so they would have identical scoring algorithms. For example, if an item was scored 0-1 in 2006-2007, but 0-1-2 in 2007-2009, the difficulty and step values would be anchored only to the 0-1 values for the 2007-2009 analysis. The model would then not expect students to score a 2 on the item, given that the maximum score on the anchored value is a 1. If these steps had not been employed, estimates on all items would have become skewed.

Occasionally, items had inconsistent scoring protocols. For example, most 6-value scale items in the 2007-2009 version of the test were scored 0-1-2-3-4-5, but one was instead scored 0-5-10-15-20-25. The different scoring for the one item resulted in the item being differentially weighted. All estimates in a Rasch model are based on the total sum score of the test items. If one item contributes more to the sum score, then it is weighted more heavily in the calibrations. Further, a partial credit model assumes there are no empty steps (e.g., the partial credit model would assume that 1-4 were possible score values). All items with inconsistent scoring algorithms were recoded to match the rest of the data.

Finally, on occasion there were items with missing step values. That is, while the item was intended to be scored 0-1-2, only the values of 0 and 2 were present. In these instances, items were recoded to the least condensed scale possible. For example, if an item was intended to be scored 0-1-2-3-4-5, but only had valid responses in the 0-1-2-3-5 categories, the item would be recoded to 0-1-2-3-4. Items were always recoded to have consistent scoring throughout all five years of the data. The recoding schemes of all items are reported in the equating maps (Tables 2-17 in Appendix 2.6).

Internal Review of the Items and Forms

DRA maintains iterative internal and external quality assurance procedures and review protocols designed to eliminate errors in content, grammar, and formatting, and to improve document retrieval and sharing by assigning document-naming protocols to all documents. These protocols are described in the appendix.

Appendix 2.7 ACAA 2010-11 Quality Assurance

CHAPTER 3: TEST ADMINISTRATION PROCEDURES

Overview

The Alaska Alternate Assessment is administered by trained Qualified Assessors, following a standardized scoring protocol. The assessment is administered individually to qualifying students and is scored at the time of administration by the Assessor.

Student Population Tested

This test is reserved for students with significant cognitive disabilities. Individualized Educational Program (IEP) teams make a determination whether a student is eligible to take the Alaska Alternate Assessment by following the guidelines in Alaska's Participation Guidelines for Alaska Students in State Assessments, June 2011 edition, located on pages 26-27 at:

http://www.eed.state.ak.us/tls/assessment/pdf_files/ParticipationGuidelinesWeb_2011.pdf

Standard Administration With or Without Accommodations

The Alaska Alternate Assessments in reading, writing, mathematics, and science are comprised of Standard test items and Expanded Levels of Support (ELoS) test items. The standard test administration uses standardized test items, student materials, and delivery instructions. The ELoS test items offer increased support and flexibility. The ELoS items are available for students who meet the criteria that are explained below. Every year, ALL students who are eligible for the Alaska Alternate Assessment must begin with the administration of the standard test tasks and items for the student's grade level. The students may use accommodations/assistive technology during testing.

Grade Level Assessments

Alaska Alternate Assessments for reading, writing, and mathematics are administered in grade bands: students in grades 3 and 4 take the 3/4 tests; students in grades 5 and 6 take the 5/6 tests; students in grades 7 and 8 take the 7/8 tests; and students in grades 9 and 10 take the 9/10 test. The Alaska Alternate Assessment in science is administered in grades 4, 8, and 10. Selecting the correct grade level assessment is critical as the scores for students testing in the incorrect grade level are invalidated. For students on the non-diploma alternate assessment track, there are no tests administered after grade 10. The Alaska Alternate Assessment is the alternate assessment for both the Standards Based Assessments and the Terra Nova.

Including Student Participation and Performance

Students taking the Alaska Alternate Assessments (including students who take the ELoS items) are counted in their school and district for Adequate Yearly Progress (AYP) in the areas of performance and participation. Individual student scores are calculated and assigned a proficiency level. The proficiency levels are: Advanced, Proficient, Below

Proficient, and Far Below Proficient. The ELOS items receive scores, but the proficiency level is Far Below Proficient as the items are non-standardized. All students receive an individual student report.

Standard Test Administration

The intent of administering the standard test items first is to provide an opportunity for each student to show what they know and can do in the grade level skills reflected in the standard administration of the Alaska Alternate Assessment. However, if a student is non-responsive, refuses to answer, or consistently earns zero scores (following the three-task, three-item rule described below), the standard administration should be stopped and the assessor must administer the Expanded Levels of Support (ELOS) test items. The purpose of stopping the standard test administration is to avoid having to administer the entire test to students who are not yet able to demonstrate skills at that level.

Standard Test Administration with Accommodations

The Alaska Alternate Assessment allows for accommodations to be utilized during test administration. Accommodations for the student are determined by the student's IEP team. The Participation Guidelines recommends that an accommodation should be used in the classroom for at least three months prior to testing. This timeline is a suggestion. It is important that the student have practice with the accommodation prior to testing; how much practice will differ by student. This amount of time allows the student to become familiar with the accommodation and ensures that the accommodation is appropriate for the student.

Standard Administration With or Without Accommodations AND Then Switched to the ELOS

The purpose of ELOS items is to provide access to the grade level tests for all students, even those who struggle with the standard alternate assessment test items. The focus of the ELOS is on students who have very limited or emerging systems of communication (e.g., may look at a speaker when her name is called, may indicate choice between activities, may have very early pre-skills for academic areas, etc.).

In each content area the Assessor must administer a minimum of three tasks and three items within each task. For each of the minimum three tasks, the student must be presented with at least three items in the task before moving on to the next task. When the student scores zeros on three consecutive items in three consecutive tasks, the Assessor should stop the assessment for that content area and must administer the required number of ELOS test items.

The three task-three item rule is operationalized as follows:

Start with Task 1 of the standard administration of the alternate assessment and proceed with successive tasks. Generally, the early tasks in each content area are easier, and tasks become progressively more difficult.

- Task 1-The assessor engages the student with the first item on a task and enters a score of zero if the student has (a) no interactive behaviors or no response, (b) actively refuses to engage in the activity, or (c) gives an incorrect answer. Next, the assessor presents the second item and enters a score of zero if the student has (a) no interactive behaviors or no response, (b) actively refuses to engage in the activity, or (c) gives an incorrect answer. Finally, the assessor moves to the third item and enters a score of zero if there is no response, the student refuses, or the student gives an incorrect answer.
- Task 2-The assessor then administers the next set of items and enters a score of zero if again there is no response, the student refuses, or the student gives an incorrect answer. When there are zeros for three consecutive items in task two, the assessor stops administering items in this task and moves to the next task.
- Task 3-Finally, the Assessor administers the next set of items and enters a score of zero if again there is no response, the student refuses, or the student gives an incorrect answer. When there are zeros for three consecutive items in task three, the assessor stops administering items in this task, and the Assessor stops the standard assessment in this content area. The Assessor must now administer the ELOS items in this content area. ELOS items may be administered immediately to complete the assessment for this content area, or at a later time.
- When a Task or Tasks have fewer than three items, Assessors are instructed to interpret the 3 X 3 rule to mean "nine consecutive zeros across a minimum of three tasks."

ELOS Administration

The ELOS test items progress from simple to more difficult items within each of three tasks. Each ELOS task has five items. Assessors must present all fifteen items to the student. Students are scored based on the degree of independence with which they complete each item:

ELOS Scores

- 1- Full physical contact to elicit student response
- 2- Partial physical contact to elicit student response
- 3- Visual, Verbal, and/or Gestural Prompts to elicit student response
- 4- Independent: No contact and no prompting needed to elicit student response

Accommodations

The Alaska Alternate Assessment allows accommodations to be utilized during test administration. Accommodations for each student are determined by the student's IEP team.

Accommodations fall into the following categories:

- **Timing/Scheduling** (e.g., extended time, frequent breaks, etc.)
- **Setting** (e.g., study carrel, student's home, separate room, etc.)
- **Presentation** (e.g., repeat directions, read aloud, large print, Braille, etc.)
 - Included with Presentation is **Assistive Devices/Supports** (e.g., calculator, amplification equipment, manipulatives, etc.)
- **Response** (e.g., mark answers in book, scribe records response, point, use an assistive device, etc.)

The Participation Guidelines recommends that an accommodation should be used in the classroom for at least three months prior to testing. This timeline is a suggestion. It is important that the student have practice with the accommodation prior to testing; how much practice will differ by student. This amount of time allows the student to become familiar with the accommodation and ensures that the accommodation is appropriate for the student. A participation guideline is available on the EED website at: http://www.eed.state.ak.us/tls/assessment/pdf_files/ParticipationGuidelinesWeb_2011.pdf

Test Administrators

Only school personnel may administer the Alaska Alternate Assessment. This includes both teachers and paraprofessionals. In order to become a Qualified Assessor (QA), individuals must participate in online training, pass proficiency tests, and administer a practice assessment that is then reviewed by their Qualified Mentor-Trainer (QT). Each QT must go through this same training, as well as additional in-person training provided annually by the EED and DRA, in order to serve as a valuable resource to QAs. These individuals have been appointed by the Special Education Director or Superintendent to be the primary point of contact for EED's Alternate Assessment Program Manager.

Mentor Responsibilities

A district appoints a person to become a QT. A Mentor-in-training first must meet all of the training requirements to become certified as a QA.

A mentor-in-training attends new mentor training as well as the annual mentor training. Both trainings are provided by EED and include:

- Complete all required training
- Receive materials to support training (PPT, handouts, examples of scoring protocols)
- Train a protégé to become a QA by:
 - Providing orientation to assessments and online training program and ongoing support
 - Reviewing and providing feedback to protégé on practice tests after they achieve proficiency on the online training
 - Upgrading protégé status from AIT to QA, after protégé have produced corrected scoring protocols to the qualifying level

After meeting qualifications, QTs become certified and have their status upgraded by EED or DRA.

Ongoing requirements to continue as a QT:

- Hold a QT Certificate, or attend new QT training
- Attend any required refresher trainings
- Refresh proficiency annually to maintain access to online system
- Sign Test Security Agreements annually and give to EED with a copy on file with the District Test Coordinators (DTCs)

Mentors have access to online reports to track their protégés' progress through training, update their status to QA when appropriate, track progress toward entering student demographic information, progress toward completion of assessment administrations, and to track any Assessors who have not completed student assessments during the last week(s) of the testing window.

Materials

All materials used in training are available to QTs for use in their respective districts to train and certify their new QAs. Materials are organized into sections on the ak.k12test.com website. Some material is restricted to personnel with QT status and higher, secure test documents are restricted to personnel with QA status or higher. The training pages and support materials for training are available to all registered users.

Test Administrator Training

Special education teachers who were selected by their districts to serve as QTs for the Alaska Alternate Assessment attended a two-day, New Mentor training in Juneau on September 27 and 28, 2011. After these Assessors-In-Training (AITs) completed all training and proficiency tests successfully, they administered a practice test, which was reviewed by DRA. Once the AIT completed these tasks, his or her account was updated to the status of QA. During training, these participants also scored a protégé's assessment protocols. After passing all these tasks, participants were upgraded to QT status, and were invited to attend the All Mentor Training in October 2011.

The additional responsibilities of a QT necessitate additional training, which was held October 25-26, 2011 in Anchorage. This training provided more in-depth information on the creation of and changes to the 2010-2011 Alaska Alternate Assessments and Secure website, including training tips to the QTs.

New Mentor Training

The purpose of the Alaska Alternate Assessment Mentor Program is to prepare district level trainers who train district personnel in correct test administration procedures for the Alaska Alternate Assessment. Mentors are available throughout the year to answer questions and assist district personnel. They are the first point of contact in the district for EED's Alternate Assessment Program Manager. Additionally, Mentors act as an advisory group for the Alaska Alternate Assessment. Mentors should be certified teachers in the State of Alaska with a special education endorsement and have experience with low-incidence disabilities. The state encourages every district to have at least one QT and one QA.

The bulk of training occurs on the website <http://ak.k12test.com>. AITs participate in a series of video vignettes designed to familiarize them with both appropriate testing and scoring techniques. These training vignettes familiarize AITs with the wide variety of tasks they will encounter on the Alaska Alternate Assessment, and demonstrate all the nuances needed in a proper administration. Following the training exercises, AITs must pass a series of brief proficiency tests related to the different tasks in each content area, as well as tests on general administration.

Summary of Dates and Participants

The New Mentor training was conducted in Juneau, AK, on September 26 and 27, 2011. After a brief introduction, instruction was given in obtaining passwords and login identities, and navigating through the Alaska Alternate Assessment training and score entry website. Participants then completed online training and proficiency tests for each of five content areas: test administration, reading, writing, mathematics, and science.

Appendix 3.1a New Mentor Training Attendees

Appendix 3.1b New Mentor Training Agenda

Appendix 3.1c New Mentor Training Handouts

Scoring Accuracy Analysis

The second day of training was devoted to gaining proficiency in administering the test to a "protégé." Participants administered and scored practice tests to each other. Participants also rated a protégé's assessment in reading, writing, mathematics, science, and ELOS. The protégé assessment that was scored was a fabricated set of assessments; all AITs received the same set of scored assessments. This allows for consistent scoring across Assessors, as well as tracking scoring issues across training years.

Annual Mentor Training

Annual Mentor training was held October 25 and 26, 2011 in Anchorage, AK. Participants from the New Mentor Training in September attended as well as veteran QTs.

Appendix 3.2a Annual Mentor Training Attendees

Appendix 3.2b Annual Mentor Training Agenda

Appendix 3.2c Annual Mentor Training Handouts

A new *Writing Scoring Manual* was introduced, as well as a guided practice module and an independent practice module for use in both training QTs and for QTs to use in their respective trainings of QAs and AITs. During the guided practice module, the *Writing Scoring Manual* was reviewed and discussed, including rules for scoring Correct Word Sequences, Correct Letter Sequences, Ideas and Organization, and Legibility, along with multiple examples of different applications. This review was coordinated with the *Writing Scoring Mini-lesson* PowerPoint, Mini-lesson workbook, and Mini-lesson answers document (which was reviewed with participants after the activity). Participants also completed a writing scoring workbook, the independent practice module, after reviewing the manual and multiple examples as a whole group. The writing scoring workbook, called the *Writing Scoring Activities*, was conducted in small groups and presented scoring challenges to participants for every rule presented in the *Writing Scoring Manual*. The workbook includes several strategies to use when training adults, as defined by Speck (1996). These adult engagement strategies were explained and modeled for participants as part of the train-the-trainer model. Participants generated written products that their partners had to score, trying to utilize examples that were relevant to the challenges they discover in the field. The challenge was to generate a written product that violated only the rule in question. This allowed for many teachable moments, both in generating the product and in scoring the outcomes. Please see Appendix 2.4 for the training documents.

Webinar

On January 26, 2012, DRA and EED hosted a web-based seminar (webinar) for QTs of the updates to the Alaska Alternate Assessment website. The webinar served to update QTs to improvements to the Alaska Alternate Assessment website (led by DRA) and to update them on procedural information related to the Alaska Alternate Assessment system and procedures (led by EED).

Appendix 3.3 Webinar

Online Training

All Assessors must complete the online training through the ak.k12test.com site. After completing training, Assessors complete proficiency testing. After participating in training through the ak.k12test site on all aspects of administering, scoring, and data entry for the Alaska Alternate Assessment, Assessors-in-Training (AITs) participate in proficiency testing. Each of five training areas are tested with a 20-question multiple-choice test (Administration, Reading, Writing, Mathematics, and Science). AITs are given two opportunities to earn a passing score of 80% or greater. If the AIT is unsuccessful in two attempts, the AIT must contact his or her Qualified Mentor to reset the proficiency tests. The AIT then has another two opportunities to pass the test in that specific domain.

Appendix 3.4 Training Site Table of Contents

Qualified Mentors are encouraged to analyze the AIT's performance on the proficiency assessment and compare that to other data available through the ak.k12test.com Web report function.

Appendix 3.5 Website Report Specifications

Refresher Training and Testing

Returning QAs and returning QTs who completed training in 2009-2010 were eligible to participate in a more efficient training and a refresher-proficiency test. After completing the reduced training sections, returning Qualified Assessors and Qualified Mentor-Trainers completed a 25-question multiple-choice Refresher Proficiency Test. All five areas of training (Administration, Reading, Writing, Mathematics and Science) are assessed, and a score of 80% is required for passing.

After two failed attempts at passing the Refresher Proficiency Test, a returning Qualified Assessor's or Qualified Mentor-Trainer's test sessions are reset to the full set of five proficiency tests.

- Returning Qualified Assessors were instructed to contact their Mentor to reset the refresher tests.
- Returning Qualified Mentor-Trainers were instructed to contact EED to reset the refresher tests.

Appendix 3.6 Refresher Training Tasks

Security

Items and test documents are maintained in a secure fashion. Transfer of items or documents containing secure test items or documents containing FERPA-protected student information are made via a secure file transfer site.

During annual training, all participants are required to sign and return a test security agreement. This document reiterates the message from training: test security is of the utmost importance in obtaining valid and reliable scores. As such, QAs must keep all materials in a confidential location, and refrain from discussing specifics of the test with others. Following the close of the test administration window, all testing materials should be shredded (with the possible exception of test documents that are used by EED for relevant studies). Teachers cannot access the secure test documents until they have passed the training requirements (passing all proficiency tests and, for Assessors-In-Training, administration and submission of a practice test). After completion of all requirements, they are granted access to the secure test materials.

The Test Security Agreement is available in the appendix.

Appendix 3.7 Test Security Agreement

The ak.k12test.com and akreports.k12test.com websites are maintained in a secure and protected system, detailed in the appendix.

Appendix 3.8 Test Site Security

CHAPTER 4: SCORING

Overview

All Qualified Assessors complete the entire online training and proficiency testing. To become a QT, QAs participate in additional training, including administering and scoring a practice test, and reviewing an assessment and scoring procedures of a protégé. These tools were analyzed to determine efficacy of training around scoring. The protégé tool is not included in the appendices, as this tool is used each year.

Quality Control of Scoring

Procedures

Alaska educators who are new to administering the Alternate Assessment and Qualified Assessors who desire to become Qualified Mentor-Trainers (new mentors), are trained at the New Mentor Training. They complete a rigorous online training protocol (described in the previous section). At the conclusion of online training, new Assessors and new Mentors complete an online proficiency test. Participants must earn 80% or higher in each subject area (administration, Reading, Writing, Math and Science) and have two opportunities to do so. If the Assessor does not earn the required proficiency within two trials, he or she must contact a Mentor to have the trials reset for additional attempts.

DRA completes an analysis each year on the number of trials required to reach proficiency in the online proficiency tests. The ak.k12test.com site collects data as users access every tool available. A review of the number of Assessors who passed a given subject area's proficiency test (with the total number of Assessors who attempted the test) is shown below. Raw Assessor proficiency data has been shared with EED, but is not reported here due to teacher confidentiality.

| Assessment | 1 st attempt | 2 nd attempt | 3 rd attempts | 4 th attempt | 5 or more attempts |
|----------------|-------------------------|-------------------------|--------------------------|-------------------------|--------------------|
| Administration | 100 of 107 | 2 of 4 | 1 of 1 | NA | NA |
| Reading | 59 of 107 | 19 of 28 | 3 of 4 | 1 of 1 | NA |
| Writing | 49 of 105 | 17 of 44 | 13 of 20 | 2 of 6 | 1 |
| Math | 75 of 106 | 5 of 16 | 5 of 10 | 0 of 5 | 5 |
| Science | 100 of 105 | 3 of 3 | NA | NA | NA |
| Refresher | 186 of 195 | 2 of 4 | 0 of 1 | NA | NA |

In addition, an analysis of the Proficiency and Refresher test item performance highlighted several test items that bear further analysis for complexity.

Appendix 4.1 Proficiency Test Item Analysis

Scoring Practice tests: The second day of New Mentor training was devoted to gaining proficiency in administering the test to a “protégé,” as described in the previous section. Participants demonstrated scoring accuracy ranging from 98% accurate in reading to 87% accurate in science, with an overall accuracy of 93% across all four content areas and ELOS items.

Scoring Protégé Reviews: In terms of reviewing protégés, participants demonstrated scoring accuracy ranging from 100% in writing and ELOS to 89% in science, with an overall accuracy rating of 93% in administration and scoring of protégés.

Qualified Mentors are responsible for training new Assessors within their districts, using the online training, proficiency testing, and practice test administration protocols established at New Mentor Training.

Appendix 4.2 New Mentor Inter-rater Reliability Report

Appendix 4.3a-4.3m Practice Tests

Inter-rater Agreement

Reliability in scoring is obtained through required intensive training online, and in administering practice tests that are reviewed by a Qualified Mentor. These steps are detailed in Chapter 3.

Handling of Exceptional Cases

Two participants in New Mentor Training required additional coaching to pass the task of assessing a protégé’s test administration. Both participants eventually met required expectations and completed their training successfully.

Writing Study

Student materials and scoring protocols for students in grades 8 and 10 were collected from District Test Coordinators statewide from April 13-20, 2012. These materials were analyzed based upon Correct Word Sequences scoring, Ideas and Organization scoring, and Legibility scoring. Expert raters reviewed all materials and provided scores for these measures for all students with viable records. These ratings were then compared to the Assessor's ratings for inter-rater reliability statistics.

The results from this year's writing scoring accuracy study revealed that the writing results demonstrate an acceptable degree of reliability.

Descriptive statistics reveal large variations among the student population, with standard deviations approaching or exceeding mean scores in many cases. The descriptive statistics also demonstrate consistent ratings among the three raters (R1 = Alaska Assessor/ R2 & R3 = Expert Rater). The ratings hang together, with means and standard deviations within each rater grouping reflecting largely consistent results.

Hierarchical Linear Modeling (HLM) analyses showed that only .06%/.03% of the item variance was due to the rater, while 57%/34% was due to the item (Grade 8/ Grade 10). This is consistent with what is expected and connotes an assessment in which rater interference is minimal. Analysis of the cross-tabulation tables for both Grade 8 and Grade 10 demonstrates that item scoring was largely consistent across raters, with lack of agreement in the majority of cases being within one unit.

High Intraclass Correlation Coefficients (ICCs) indicate stable measurement on the writing assessments. The results of the analysis reflect positively upon the reliability of the writing assessments, with ICCs ranging from .907 to .997 in Grade 8, and from .941 to .999 in grade 10.

Cross-tabulation tables demonstrate a high degree of scoring accuracy, with most discrepancies limited to one unit above or below consensus ratings. Kendall-tau-c statistic results range from .468 in R1/R3 Legibility raw score ratings to .944 in R2/R3 CWS points ratings in Grade 8. In Grade 10, the Kendall-tau-c statistic results range from .308 in R1/R3 raw score ratings in Ideas and Organization to .996 in R2/R3 ratings in CWS raw points possible. Thus, additional fortification of accuracy in rating Legibility in Grade 8 and Ideas and Organization in Grade 10 is needed. The Kappa statistics for Grade 8 range from .452 comparing R1/R3 on CWS_{pct} to .951 comparing R2/R3 on CWS_r. Results thus reflect moderate to almost perfect agreement in Grade 8. The Kappa statistics for Grade 10 range from .211 comparing R1/R3 on IO_r to 1.000 comparing R2/R3 on rpp and IO_r. Results thus reflect fair to almost perfect agreement in Grade 10. The Kappa statistics should also be compared to the cross-tabulation tables, as small deviations in scoring can occasionally lead to large differences in agreement. Grade 10 results are presented in Tables 32-45. Grade 10 includes the variables mentioned above for Grade 8 as well as rpp (raw points possible in CWS).

Appendix 4.4 Writing Scoring Study Documents

Data Entry

After entering each student eligible for an Alaska Alternate Assessment on their caseload to the online system, assessors enter student scores into the ak.k12test.com site, on the Data Entry page.

The student's grade of enrollment preloads the possible assessments available for that student. Assessors enter the scores for each item in each eligible assessment, or indicate a reason not tested.

After entering scores in all available subject areas, Assessors are prompted to submit the scores to EED. There are two ways to submit scores to EED.

1. After all scores for all required assessments have been entered, the system prompts the QA to submit the data to EED. QAs may select this option to "Submit" the data at this point; or,

2. Alternately, a QA may return to the Data Entry page and mark the record as complete by choosing the appropriate status in the Status of Data Entry drop-down box in the left-hand column.

To mark the record complete, the Assessor must have entered data for each subject or given a reason why the test wasn't administered. If a subject area assessment is not administered for a student, the Assessor must choose a "Reason Not Tested" for that assessment. Scores not submitted by the close of the testing window are invalidated.

Following are the "reasons not tested" that a QA or DTC would choose to alert DRA and EED about why they are not testing a student. This information is located in the Data Entry section of the online assessment system. Students may participate in one or more Alaska Alternate Assessment content areas, and may not be eligible to participate in the AA-AAs in the other content areas tested.

1. IEP Change This code is selected for students who have an IEP change indicating they are no longer eligible to take the Alaska Alternate Assessment in one or more content areas, and will be taking the Standards Based Assessment (SBA) instead. This code should only be selected for the content areas in which the student is not taking the Alaska Alternate Assessment.

2. Late Entry This code is entered for students who enter the district from out of state or from a private school after the Alaska Alternate Assessment test window opens. In order to count for the district's participation rate, the district must administer a minimum of one assessment in reading, writing, or mathematics. This code should only be selected for the content areas in which the student is not being assessed.

3. Suspension The student is suspended or expelled for the entire test window. If this code is selected, it automatically applies to all content areas.

4. Other Any other reason must be documented in a text box that will appear when the "Other" code is selected. This code should only be selected for the content areas in which the student is not assessed. Text is limited to 50 characters, including spaces.

Beginning with the 2010-2011 testing window, EED lengthened the testing window to 10 weeks. "Long Term Absence" is no longer an approved reason for not testing a student in the Alaska Alternate Assessment.

For the 2011-2012 assessment, as in 2010-2011, Assessors were asked to define the scheduling of test administration during the data entry process. Assessors chose one of four options:

Timing/Scheduling Accommodations:

- A. This subject administered with breaks/multiple sessions
- B. This subject administered with NO breaks/one session

C. Multiple-subject administration with breaks/multiple sessions

D. Multiple-subject administration with NO breaks/one session

Choose all tests administered in one session: [checkboxes, allow one or all to be chosen]

0 Reading 0 Writing 0 Math 0 Science

"Breaks" means that the student was provided frequent breaks during testing.

"Multiple sessions" means that the test was administered over several days.

Assessors employed timing and scheduling accommodations in the following amounts:

| Choice | Reading | Writing | Math | Science |
|--------|---------|---------|------|---------|
| A | 206 | 185 | 203 | 61 |
| B | 255 | 273 | 260 | 125 |
| C | 135 | 136 | 129 | 33 |
| D | 62 | 61 | 60 | 23 |

In addition, teachers of students eligible for the Alaska Alternate Assessment are encouraged to use the Alaska Alternate Assessment practice tests throughout the school year with their students. The practice tests allow the teacher/test administrator to become more comfortable manipulating the testing materials, allow the teacher to test the efficacy of accommodations with students in testing situations, and allow the teacher to develop an understanding of student stamina and tolerance for performance testing tasks. In addition, the teacher may help the student develop test-taking strategies and become comfortable with the Alaska Alternate Assessment testing format prior to administration of the official Alaska Alternate Assessment.

There is a certain amount of flexibility for the test administrator with regard to how to present student materials. In addition to altering the materials for an allowable accommodation (e.g., increasing the text size of student materials), real-life objects may be substituted for those represented in the materials. For example, an actual glass of water may be used in lieu of the drawing of a glass of water provided in the materials, if this makes the test item more accessible to the particular student. Large Print and Braille tests are also available.

The QA may position himself in any location that is most helpful for managing the assessment materials, the student's behaviors and access to the assessment materials, and the scoring protocols. This may be side by side with the student, across the table from the student, or any position that works for the assessor and the student.

Additional resources on accommodations are available:

EED Accommodations website:

<http://www.eed.state.ak.us/tls/assessment/accommodations.html>

National Center on Educational Outcomes accommodations website:

<http://www.cehd.umn.edu/NCEO/TopicAreas/Accommodations/Accomtopic.htm>

Analysis of Accommodations Used

DRA collected data from Assessor input regarding the accommodations used in the 2011-2012 assessments. For each subject area, the total number of accommodations made is reported, as well as the total number of students receiving one or more accommodations to the Alaska Alternate Assessment.

- In Reading, there were 28 types of accommodations made to the test administration with 8 types of accommodations made for student responses and 3 types of accommodations made to the test materials. A total of 118 students utilized one or more type of accommodation in Reading.
- In Writing, there were 30 types of accommodations made to the test administration with 10 types of accommodations made for student responses and 6 types of accommodations made to the test materials. A total of 106 students utilized one or more type of accommodation in Writing.
- In Mathematics, there were 29 types of accommodations made to the test administration with 2 types of accommodations made for student responses and 2 types of accommodations made to the test materials. A total of 142 students utilized one or more type of accommodation in Mathematics.
- In Science, there were 18 types of accommodations made to the test administration with 4 types of accommodations made for student responses and 2 types of accommodations made to the test materials. A total of 41 students utilized one or more type of accommodation in Science.

In 2010-2011, some accommodations appeared to cross the threshold into modifications (the adaptation could have changed the construct being tested or provided too much assistance to the child in the testing setting). The Mentors in three districts were notified, and the accommodation in question was described. The Mentors worked with the Assessors to determine the appropriateness of the accommodation.

For the 2011-2012 testing year, the number of different types of accommodations that might be modifications were reduced. "Hand over hand assistance" is a type of accommodation that appeared frequently, across all subject areas (3 times in Reading, 20 in Writing; 3 in Math and 2 times in Science). Accommodations that are possible modifications are marked with an asterisk (*) in appendix 4.5. Accommodations which were marked with an asterisk in 2010-2011, but which had zero instances in 2011-2012 are left in the list to show improvement in this area. Raw accommodations data has been shared with EED, but is not reported here due to student confidentiality.

Appendix 4.5 Accommodations Used Summary

Scoring ELOS Tasks

ELOS tasks are scored one through four. Scores are defined in the Levels of Independence Scoring Rubric. The additional levels of support are designed to bring the student to success. The Assessor begins with the least amount of additional support (e.g., the Assessor asks the question and waits for the student to respond), and introduces successively greater amounts of support, as needed by the student. Drawing the student's attention to the page by pointing in general to the answer choices is not considered a gestural support. A gestural support in ELOS is when the Assessor points to the correct answer ("Which one is the math problem" -- "This one (pointing to the math problem) is the math problem. Can you point to the math problem?").

ELOS Scores

- 1- Full physical contact to elicit student response
- 2- Partial physical contact to elicit student response
- 3- Visual, Verbal, and/or Gestural Prompts to elicit student response
- 4- Independent: No contact and no prompting needed to elicit student response

CHAPTER 5: STANDARDS VALIDATION

The Alaska Alternate Assessment did not undergo a standards validation analysis this year.

(This page left blank by design)

CHAPTER 6: REPORTING

Overview

A number of tables are presented in the appendix, displaying various statistics for use in interpreting the Alaska Alternate Assessment reports.

All tables and analyses are presented for subject area results in a standardized layout format: reading, writing, mathematics, and science.

Frequency counts are used to display the number and percentage of students at various grade bands. The number and percentage of students at each score value are also displayed in the appendix. An important statistic in every table is the valid N or the number of students represented in the statistic for any given measure. Means and standard deviations are used to describe the distributions at various grade bands. These two statistics should be interpreted relative to each other; ideally, the Standard Deviation (SD) is less than (even half) the amount mean, which can be interpreted as reflecting an appropriate amount of variation. When the SD is close to or greater than the mean, then the distribution is difficult to describe as there appears to be as much variation as there is centeredness. Minimums and maximums reflect the smallest and largest scores obtained on the test, respectively.

Many tables have a total that simply reflects the sum of any frequency count across all categories (e.g. grade level or score value). System missing refers to the number of students who are not in that statistical calculation (either frequency or mean).

When reliability coefficients are displayed, a value is presented that varies from a low moderate decimal (in the .30-.50 range), a moderate range (.51 to .79) or a relatively high value (in the .80 to .97 range). These values represent the degree to which two variables (e.g. forms of the test or items within the test) are related. Generally, higher is better, as the information from one measure (item or form) can be used to predict another item or form. In some cases, however, the values should not be too high (e.g., when reflecting the relations among different items in the test), because it would mean that, essentially, they are duplicating the information.

This statistic, however, is a function of the number of values (in the test) that are counted (as well as the number students behind any of these values). For example, at the total test level, many items are used to calculate the coefficient; at the strand level, sufficient items are present. However, at the task level, the number of items is so few that the values are likely to be low because there simply is not enough variation present to reflect a high coefficient.

Reporting Student Results

Two score reports are generated for each student: an Unofficial Score Report and an Official Score Report. The **Unofficial Score Report** is generated immediately on completion and submission of student scores for all eligible alternate assessments. This report is an exact accounting of the student's performance. **Official Student Reports** are released to the District Test Coordinator in mid-May, after the AYP calculations are completed. The Official Student Report reports a student's proficiency level relative to the Extended Grade Level Expectations (ExGLE).

Chapter nine fully describes the calculations, results, and reporting methodologies for AYP.

The differences between these two sets of scores are explained in a comparison chart, available in the appendix.

Appendix 6.1 Unofficial and Official Individual Student Report Matrix

The appendix also lists sample documents used in reporting student results, including an Unofficial Student Report, and Official Student Reports in Reading, Writing, Math, and Science, and Guides to educators and parents on reading and understanding student score reports.

Appendix 6.2a Educator Reading, Writing, Math and Science Report Examples

Appendix 6.2b Parent Reading, Writing, Math and Science Report Examples

DRA Secure Reporting Website

Official Individual Student Reports were made available to each district's District Test Coordinator on May 16, 2012. Beginning this year, Qualified Mentor-Trainers also received passwords to access the secure reporting site. Reports are downloaded from the secure Reporting Website at akreports.k12test.com. District Test Coordinators are given a secure user identification and password at an annual training, held in February.

Reports are bundled for each district by school and then by student last name.

Appendix 6.3 Reporting Website Manual

CHAPTER 7: TEST VALIDITY

Overview

The statistical data output for Chapter 7 and Chapter 9 are located in the appendix in the folder for Chapter 7. The data output for Chapter 8, strand, task, and item difficulty statistics, are included in the body of the technical report. The document for each subject area contains the output regarding AYP calculations, test strand descriptive statistics, task descriptive statistics, task item descriptive statistics and reliability statistics.

Validity

As elaborated by Messick (1989)¹, the validity argument involves a claim with evidence evaluated to make a judgment. Three essential components of assessment systems are necessary: (a) constructs (what to measure), (b) the assessment instruments and processes (approaches to measurement), and (c) use of the test results (for specific populations). To put it simply, validation is a judgment call on the degree to which each of these components is clearly defined and adequately implemented.

Validity is a unitary concept with multifaceted processes of reasoning about a desired interpretation of test scores and subsequent uses of these test scores. In this process, we want answers for two important questions. Regardless of whether the students tested have disabilities, the questions are identical: (1) How valid is our interpretation of a student's test score? and, (2) How valid is it to use these scores in an accountability system? Validity evidence may be documented at both the item and total test levels. We use the *Standards*² (AERA et al., 1999) in documenting evidence on content coverage, response processes, internal structure, and relations to other variables. This document follows the essential data requirements of the federal government as needed in the peer review.³ The critical elements highlighted in that document (with examples of acceptable evidence) include (a) academic content standards, (b) academic achievement standards, (c) a statewide assessment system, (d) validity, (e) reliability, and (f) other dimensions of technical quality.

This document addresses the latter four requirements (c-f noted above), with other documents providing essential information on the standards and statewide assessment system (see technical specifications and alignment documents for information on academic content standards and the standard setting document for information on the academic achievement standards). In addressing technical documentation, we first present content evidence, then reliability, and finally address the other three areas noted in the peer review guidance: response process, internal structures, and criterion relations.

¹ Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education.

² American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

³ U. S. Department of Education (2004). *Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Left Behind Act of 2001*

Content related evidence includes information on technical specifications and the quality of review used during the design and development of the alternate assessment. In particular, we emphasized ‘universal design’ in developing items and tasks that would be clear enough in their presentation and sufficiently flexible in their administration to allow ALL students access. This outcome was achieved through both the item writing and reviewing in which content experts and special educators provided feedback through the stages of test development. We also summarize outcome data as a reference for understanding subsequent validity evidence for content skills and knowledge.

Reliability

The data file was analyzed for reliability at several levels. First, at the total test level, which is the most important because Adequate Yearly Progress (AYP) is established on the basis of this score, reliability coefficients are reported for every grade band and subject area. Second, at the strand level, coefficients are reported for every grade band and subject area. The test was designed to reflect scores at this level to ensure adequate representation across the entire range of Extended Grade Level Expectations; in the official student reports, scores for every strand are reported so that parents and teachers can follow the performance and progress of students. Third, and perhaps least important, are the scores at the task level; though we report these coefficients, they are primarily directed toward the continuous improvement of the test as EED develops new field tests and integrates them into the operational test.

In the tables for total test and strands, the reliability coefficients are reported for both the entire population (ALL students) and the students who took the complete Standard administration with students who participated in the Expanded Levels of Support (ELOS) removed (with NO ELOS). This population includes students with extremely low levels of functioning with little to no interactivity or means of communication. The reason for removing this group was to investigate the influence of missing data and its potential to spuriously inflate reliability coefficients. The first step in removing this group was to integrate the ELOS data file with the standard administration file. The second step involved splitting the file on ELOS participation and removing them so that all reliability coefficients could be recomputed at each level (total test, strand, and task). This re-analysis was done for each subject area and at all grade bands.

In general, the findings indicate that the test is very reliable for decision-making (of AYP) at the total test level. Scores were quite reliable at the strand level (with only a few strands reflecting moderate coefficients, which was primarily a function of the few number of tasks involved). Finally, as expected, scores were moderately reliable at the task level, primarily because of the few items involved. Another general (and expected) finding is that the coefficients are somewhat lower when the ELOS students scores are removed from the standard administration file although the reduction is not large, as only 9-11% of the students were administered ELOS tasks/items (see the section “Item Performance: Task Difficulty [Standard Administration, No ELOS] for summary results).

Students who participate in ELOS administration are included in the participation rate reporting for AYP; however their scores are reported as Far Below Proficient for AYP performance reporting.

Total Test Reliabilities (All Students)

Reading: Grades 3-4 (.936 for 114 students taking 39 items), grades 5-6 (.951 for 139 students taking 35 items), grades 7-8 (.949 for 112 students taking 35 items), and grades 9-10 (.927 for 136 students taking 41 items).

Writing: Grades 3-4 (.967 for 115 students taking 21 items), grades 5-6 (.906 for 146 students taking 13 items), grades 7-8 (.954 for 124 students taking 19 items), and grades 9-10 (.908 for 154 students taking 17 items).

Math: Grades 3-4 (.934 for 112 students taking 26 items), grades 5-6 (.961 for 137 students taking 51 items), grades 7-8 (.966 for 103 students taking 66 items), and grades 9-10 (.951 for 128 students taking 65 items).

Science: Grade 4 (.922 for 64 students taking 24 items), grade 8 (.908 for 67 students taking 24 items), and, grade 10 (.883 for 72 students taking 24 items).

In this analysis, all items used in a grade band **test** were entered for computing the reliability of the entire test: Cronbach's Alpha based on Standardized Items. Reliability for tasks including three or fewer items are not reported.

Appendices 7.1 – 7.4 Reading, Writing, Math, and Science Reliability Statistics

Test Task Reliabilities

Reading Reliability

Reading Grades 3 and 4

| Task Name | Cronbach's Alpha Based on Standardized Items |
|--|--|
| 1.34A: Identify Signs and Symbols | .747 |
| 1.34B Identify Letter Sounds | .900 |
| 1.34C: Blend Sounds | .948 |
| 2.34A: Read Passages: Story 1, Annie Goes to a Party | .735 |
| 2.34B: Read Passages: Story 2, Jill and the Zoo | .616 |

Reading Grades 5 and 6

| Task Name | Cronbach's Alpha Based on Standardized Items |
|--|--|
| 1.56A: Read Words | .928 |
| 1.56B: Read Sentences | .922 |
| 2.56A: Read Passages: Story 1, Jimmy Rides the Bus | .884 |
| 2.56B: Read Passages: Story 2, Jack is so Friendly | .797 |

Reading Grades 7 and 8

| Task Name | Cronbach's Alpha Based on Standardized Items |
|--|--|
| 1.78A: Read Words of Increasing Complexity | .936 |
| 1.78B: Obtain Information | .825 |
| 1.78C: Read Sentences | .960 |
| 2.78A: Read Passages: Story 1, Hannah's Homework | .877 |
| 2.78B: Read Passages: Story 2, Eating Lunch at the Cafeteria | .868 |

Reading Grades 9 and 10

| Task Name | Cronbach's Alpha Based on Standardized Items |
|--|--|
| 1.910A: Decode Words | .947 |
| 1.910B: Identify Root Words | .703 |
| 2.910A: Read Passages: Story 1, Jan and the Party | .841 |
| 2.910B: Read Passages: Story 2, Ready for Graduation | .826 |

Writing Reliability

Writing Grades 3 and 4

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---------------------|--|
| 1.34A: Copy Letters | .959 |
| 1.34B: Copy Words | .968 |

Writing Grades 5 and 6

| Task Name | Cronbach's Alpha Based on Standardized Items |
|-----------------------------------|--|
| 1.56A: Conventions of Writing | .579 |
| 1.56C: Write Words from Dictation | .939 |

Writing Grades 7 and 8

| Task Name | Cronbach's Alpha Based on Standardized Items |
|--------------------------------------|--|
| 1.78C: Communicate Ideas Using Words | .944 |
| 1.78E: Revise Sentences | .798 |

Writing Grades 9 and 10

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|--|
| 1.910A: Conventions of Standard English | .843 |
| 1.910C: Revise Writing | .841 |

Math Reliability*Math grades 3 and 4*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|-----------------------|--|
| 1.34A: Copy Numbers | .959 |
| 1.34B: First and Last | .763 |
| 2.34: Same/Different | .876 |

Math grades 5 and 6

| Task Name | Cronbach's Alpha Based on Standardized Items |
|------------------------------------|--|
| 1.56A: Read and Write Numbers | .933 |
| 1.56B: Number Line, First and Last | .830 |
| 2.56: Simple Addition | .952 |
| 3.56: Reproduce Simple Patterns | .875 |
| 4.56: Read Simple Graphs | .830 |
| 5.56B: Identify Money | .773 |
| 6.56A: Identify Shapes | .874 |

Math grades 7 and 8

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|--|
| 1.78A: Read and Write Numbers, ID Place Value | .764 |
| 1.78C: Ordering - Number Line and Pictures | .845 |
| 2.78: Double-Digit Addition and Subtraction | .917 |
| 3.78B: Label a Set as None or Zero | .934 |

| | |
|--------------------------------------|------|
| 4.78: Read Simple Graphs | .815 |
| 5.78A: Identify Units of Measurement | .812 |
| 5.78C: Identify Money | .895 |
| 6.78A: Identify Shapes and Position | .770 |

Math grades 9 and 10

| Task Name | Cronbach's Alpha Based on Standardized Items |
|---|--|
| 1.910A: Identify Place Value | .851 |
| 1.910B: Identify Fractions | .729 |
| 2.910B: Double-Digit Addition/Subtraction and Single-Digit Multiplication | .874 |
| 3.910A: Extend a Pattern/Supply Missing Element | .740 |
| 3.910B: Understand Symbols | .713 |
| 4.910: Read Simple Graphs | .832 |
| 5.910A: Identify Units of Measurement | .818 |
| 6.910A: Describe and Compare Shapes, Shapes Greater Than, Less Than, Equal To | .840 |

Science Reliability*Science grade 4*

| Task Name | Cronbach's Alpha Based on Standardized Items |
|--|--|
| 1.4: Concepts of Physical Science | .777 |
| 2.4: Concepts of Life Science | .688 |
| 3.4: Concepts of Earth Science | .783 |
| 4.4: History and Nature of Science, Science and Technology | .732 |

Science grade 8

| Task Name | Cronbach's Alpha Based on Standardized Items |
|-----------------------------------|--|
| 1.8: Concepts of Physical Science | .744 |
| 2.8: Concepts of Life Science | .751 |
| 3.8: Concepts of Earth Science | .718 |
| 4.8: Science and Technology | .693 |

Science grade 10

| Task Name | Cronbach's Alpha Based on Standardized Items |
|------------------------------------|--|
| 1.10: Concepts of Physical Science | .752 |
| 2.10: Concepts of Life Science | .809 |
| 3.10: Concepts of Earth Science | .517 |
| 4.10: Science and Technology | .657 |

CHAPTER 8: DESCRIPTIVE STATISTICS

Descriptive statistics were calculated for each task, in every subject area, and in both grade bands and grade levels. The upper right header of each page refers the reader to the type of descriptive statistics displayed. For instance, "Grade Band Total Test Descriptive Statistics" refers to the descriptive statistics at the total test level for each subject, while "Writing Task Descriptive Statistics (Grade Band 3/4)" refers to the descriptive statistics for writing at the task level, in grade band 3/4. The following statistics are reported in the tables in Appendix 7 (leftmost column to rightmost column).

Strand, Task, and Item Difficulties

Strand Difficulties in Reading, Writing, Math, and Science (Standard, No ELOS)

The tables provided below elaborate the strand difficulties for reading, writing, mathematics in grade bands 3/4, 5/6, 7/8, and 9/10. Science strand difficulties are presented in grades 5, 8, and 11. Strand names are provided, as are p -values. The p -value represents the proportion of the students responding in the keyed direction (e.g., students who received partial or full credit, with students receiving full credit contributing more significantly to the rating).

Reading Strand Difficulties

The Alaska Alternate Reading Assessment was grouped into four grade bands: Grades 3/4, Grades 5/6, Grades 7/8, and Grades 9/10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLEs), and were organized at the strand level. The most difficult strand is Form a General Understanding in Grade Band 9/10, with a 50% success rate. The easiest strand is Form a General Understanding in Grade Band 7/8, with a 70% success rate. The variable ' p ' is an index of strand difficulty, with lower numbers indicating lower rates of success.

| Strand Name | p |
|------------------------------------|------|
| Word Identification 3/4 | 0.53 |
| Form a General Understanding 3/4 | 0.56 |
| Word identification 5/6 | 0.52 |
| Form a General Understanding 5/6 | 0.59 |
| Analyze Content and Structure 5/6 | 0.58 |
| Word Identification 7/8 | 0.55 |
| Form a General Understanding 7/8 | 0.70 |
| Analyze Content and Structure 7/8 | 0.68 |
| Word Identification 9/10 | 0.62 |
| Form a General Understanding 9/10 | 0.50 |
| Analyze Content and Structure 9/10 | 0.67 |

Writing Strand Difficulties

The Alaska Alternate Writing Assessment was grouped into four grade bands: Grades 3/4, Grades 5/6, Grades 7/8, and Grades 9/10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLEs), and were organized at the strand level. The most difficult strand is Write Using a Variety of Forms in Grade Band 9/10, with a 40% success rate. The easiest strand is Revise in Grade Band 9/10, with a 68% success rate.

| Strand Name | <i>p</i> |
|--|----------|
| Write Using a Variety of Forms 3/4 | 0.57 |
| Write Using a Variety of Forms 5/6 | 0.51 |
| Structures and Conventions of Writing 5/6 | 0.63 |
| Write Using a Variety of Forms 7/8 | 0.41 |
| Structures and Conventions of Writing 7/8 | 0.50 |
| Revise 7/8 | 0.56 |
| Write Using a Variety of Forms 9/10 | 0.40 |
| Structures and Conventions of Writing 9/10 | 0.67 |
| Revise 9/10 | 0.68 |

Mathematics Strand Difficulties

The Alaska Alternate Mathematics Assessment was grouped into four grade bands: Grades 3/4, Grades 5/6, Grades 7/8, and Grades 9/10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLEs), and were organized at the strand level. The most difficult strand is Estimation and Computation in Grade Band 7/8, with a 51% success rate. The easiest strand is Geometry in Grade Band 3/4, with a 73% success rate.

| Strand Name | <i>p</i> |
|--------------------------------|----------|
| Numeration 3/4 | 0.60 |
| Functions and Relations 3/4 | 0.61 |
| Geometry 3/4 | 0.73 |
| Numeration 5/6 | 0.70 |
| Estimation and Computation 5/6 | 0.56 |
| Functions and Relations 5/6 | 0.72 |
| Statistics and Probability 5/6 | 0.65 |
| Measurement 5/6 | 0.75 |
| Geometry 5/6 | 0.66 |
| Numeration 7/8 | 0.59 |
| Estimation and Computation 7/8 | 0.51 |
| Functions and Relations 7/8 | 0.72 |

| Strand Name | <i>p</i> |
|---------------------------------|----------|
| Statistics and Probability 7/8 | 0.60 |
| Measurement 7/8 | 0.63 |
| Geometry 7/8 | 0.62 |
| Numeration 9/10 | 0.67 |
| Estimation and Computation 9/10 | 0.57 |
| Functions and Relations 9/10 | 0.69 |
| Statistics and Probability 9/10 | 0.69 |
| Measurement 9/10 | 0.66 |
| Geometry 9/10 | 0.68 |

Science Strand Difficulties

The Alaska Alternate Science Assessment was grouped into three grade level assessments: Grade 4, Grades 8, and Grade 10. Within these grade band assessments, items directly correlated to the Alaska Extended Grade Level Expectations (ExGLEs), and were organized at the strand level. The most difficult strand is Concepts of Life Science in Grade 4, with a 59% success rate. The easiest strand is Science and Technology in Grade 10, with an 84% success rate.

| Strand Name | <i>p</i> |
|---------------------------------|----------|
| Concepts of Physical Science 4 | 0.66 |
| Concepts of Life Science 4 | 0.59 |
| Concepts of Earth Science 4 | 0.71 |
| Science and Technology 4 | 0.71 |
| Concepts of Physical Science 8 | 0.69 |
| Concepts of Life Science 8 | 0.64 |
| Concepts of Earth Science 8 | 0.65 |
| Science and Technology 8 | 0.79 |
| Concepts of Physical Science 10 | 0.71 |
| Concepts of Life Science 10 | 0.77 |
| Concepts of Earth Science 10 | 0.65 |
| Science and Technology 10 | 0.84 |

Task Difficulty (Standard, No ELOS)

All task difficulty statistics for reading, writing, mathematics, and science are presented in the tables below. Task descriptions are provided, as are p -values. The p -value represents the proportion of the students responding in the keyed direction (e.g., students who received partial or full credit, with students receiving full credit contributing more significantly to the rating). The variable ' p ' is an index of task difficulty, with lower numbers indicating lower rates of success.

Reading Task Difficulties

The following tables summarize task difficulties for each grade band in reading. Reading task difficulties range from .46 to .87.

Reading Tasks Grade Band 3/4

The most difficult task in the 3/4 Grade Band is Blend Sounds, with a success rate of approximately 52%. The easiest task in the 3/4 Grade Band is Identify Own Name, with a success rate of approximately 87%.

| Task Name | p |
|----------------------------|------|
| Identify Signs and Symbols | 0.53 |
| Identify Letter Sounds | 0.58 |
| Blend Sounds | 0.52 |
| Identify Own Name | 0.87 |
| Annie Goes to a Party | 0.55 |
| Jill and the Zoo | 0.58 |

Reading Tasks Grade Band 5/6

The most difficult task in the 5/6 Grade Band is Read Sentences, with a success rate of approximately 48%. The easiest task in the 5/6 Grade Band is Jack is so Friendly, with a success rate of approximately 63%.

| Task Name | p |
|---------------------|------|
| Read Words | 0.57 |
| Read Sentences | 0.48 |
| Jimmy Rides the Bus | 0.56 |
| Jack is so Friendly | 0.63 |

Reading Tasks Grade Band 7/8

The most difficult task in the 7/8 Grade Band is Obtain Information, with a success rate of approximately 46%. The easiest task in the 7/8 Grade Band is Eating Lunch at the Cafeteria, with a success rate of approximately 73%.

| Task Name | <i>p</i> |
|-------------------------------------|----------|
| Read Words of Increasing Complexity | 0.60 |
| Obtain Information | 0.46 |
| Read Sentences | 0.56 |
| Hannah's Homework | 0.64 |
| Eating Lunch at the Cafeteria | 0.73 |

Reading Tasks Grade Band 9/10

The most difficult task in the 9/10 Grade Band is Ready for Graduation, Part 2, with a success rate of approximately 51%. The easiest task in the 9/10 Grade Band is Ready for Graduation, Part 1, with a success rate of approximately 87%.

| Task Name | <i>p</i> |
|------------------------------|----------|
| Decode Words | 0.64 |
| Identify Root Words | 0.60 |
| Follow Multi-Step Directions | 0.65 |
| Jan & the Party | 0.58 |
| Ready for Graduation: Part 1 | 0.68 |
| Ready for Graduation: Part 2 | 0.51 |

Writing Task Difficulties

The following tables summarize task difficulties for each grade band in writing. Writing task difficulties range from .37 to .72.

Writing Tasks Grade Band 3/4

The most difficult task in the 3/4 Grade Band is Copy Words, with a success rate of approximately 54%. The easiest task in the 3/4 Grade Band is Matching and Sequencing Pictures, with a success rate of approximately 61%.

| Task Name | <i>p</i> |
|----------------------------------|----------|
| Copy Letters | 0.58 |
| Copy Words | 0.54 |
| Write Own Name | 0.59 |
| Matching and Sequencing Pictures | 0.61 |

Writing Tasks Grade Band 5/6

The most difficult task in the 5/6 Grade Band is Write Words from Dictation, with a success rate of approximately 51%. The easiest task in the 5/6 Grade Band is Write Own Name, with a success rate of approximately 65%.

| Task Name | <i>p</i> |
|----------------------------|----------|
| Conventions of Writing | 0.55 |
| Write Own Name | 0.65 |
| Write Words from Dictation | 0.51 |
| Write a Sentence | 0.63 |

Writing Tasks Grade Band 7/8

The most difficult task in the 7/8 Grade Band is Write a Sentence, with a success rate of approximately 37%. The easiest task in the 7/8 Grade Band is Revise Sentences, with a success rate of approximately 72%.

| Task Name | <i>p</i> |
|---------------------------------|----------|
| Write Sentences from Dictation | 0.43 |
| Conventions of Standard English | 0.49 |
| Communicate Ideas Using Words | 0.50 |
| Write a Sentence | 0.37 |
| Revise Sentences | 0.72 |

Writing Tasks Grade Band 9/10

The most difficult task in the 9/10 Grade Band is Write a Story, with a success rate of approximately 40%. The easiest task in the 9/10 Grade Band is Revise Writing, with a success rate of approximately 68%.

| Task Name | <i>p</i> |
|---------------------------------|----------|
| Conventions of Standard English | 0.67 |
| Write a Story | 0.40 |
| Revise Writing | 0.68 |

Mathematics Task Difficulties

The following tables summarize task difficulties for each grade band in mathematics. Mathematics task difficulties range from .21 to .93.

Mathematics Tasks Grade Band 3/4

The most difficult task in the 3/4 Grade Band is First and Last, with a success rate of approximately 49%. The easiest task in the 3/4 Grade Band is Identify Shapes, with a success rate of approximately 73%.

| Task Description | <i>p</i> |
|--------------------|----------|
| Copy Numbers | .61 |
| First and Last | .49 |
| Count | .70 |
| Same and Different | .61 |
| Identify Shapes | .73 |

Mathematics Tasks Grade Band 5/6

The most difficult task in the 5/6 Grade Band is Identify Perimeter, with a success rate of approximately 36%. The easiest task in the 5/6 Grade Band is Identify Shapes, with a success rate of approximately 84%.

| Task Description | <i>p</i> |
|------------------------------|----------|
| Read & Write Numbers | .75 |
| Number Line, First & Last | .60 |
| Count Objects | .73 |
| Count | .78 |
| Simple Addition | .56 |
| Reproduce Simple Patterns | .72 |
| Read Simple Graphs | .65 |
| Same, Longer & Shorter, More | .80 |
| Identify Money | .73 |
| Identify Shapes | .84 |
| Same or Different (Shapes) | .79 |
| Identify Perimeter | .36 |

Mathematics Tasks Grade Band 7/8

The most difficult task in the 7/8 Grade Band is Identify Perimeter, with a success rate of approximately 21%. The easiest task in the 7/8 Grade Band is Match Shapes, with a success rate of approximately 93%.

| Task Description | <i>p</i> |
|--------------------------------------|----------|
| Read & Write Numbers, ID Place Value | .54 |
| Identify Fractions | .57 |
| Ordering - Number Line & Pictures | .57 |
| Identify Skip Patterns | .60 |
| Count | .73 |
| Double Digit Addition & Subtraction | .51 |
| Reproduce & Extend Simple Patterns | .70 |
| Label a Set as None or Zero | .73 |
| Understand Symbols | .83 |
| Read Simple Graphs | .60 |
| Identify Units of Measurement | .59 |
| Count Money | .49 |
| Identify Money | .88 |
| Identify Shapes & Position | .83 |
| Match Shapes | .93 |
| Identify Perimeter | .21 |

Mathematics Tasks Grade Band 9/10

The most difficult task in the 9/10 Grade Band is Identify Perimeter, with a success rate of approximately 33%. The easiest task in the 9/10 Grade Band is Describe & Compare Shapes Greater Than, Less Than, Equal To, with a success rate of approximately 83%.

| Task Description | <i>p</i> |
|---|----------|
| Identify Place Value | .62 |
| Identify Fractions | .63 |
| Order Numbers | .75 |
| Round Numbers | .47 |
| Double-Digit Addition/Subtraction/Multiplication | .62 |
| Extend a Pattern/Supply Missing Element | .74 |
| Understand Symbols | .59 |
| Read Simple Graphs | .69 |
| Identify Units of Measurement | .72 |
| Count Money | .60 |
| Describe & Compare Shapes Greater Than, Less Than, Equal To | .83 |
| Lines of Symmetry | .61 |
| Identify Perimeter | .33 |

Science Task Difficulties

The following tables summarize task difficulties for each grade band in science. Science task difficulties range from .59 to .84.

Science Tasks Grade 4

The most difficult task in Grade 4 is Concepts of Life Science, with a success rate of approximately 59%. The easiest task in Grade 4 is Concepts of Earth Science and Science and Technology, with a success rate of approximately 71%.

| Task Name | <i>p</i> |
|------------------------------|----------|
| Concepts of Physical Science | 0.66 |
| Concepts of Life Science | 0.59 |
| Concepts of Earth Science | 0.71 |
| Science and Technology | 0.71 |

Science Tasks Grade 8

The most difficult task in Grade 8 is Concepts of Life Science, with a success rate of approximately 64%. The easiest task in Grade 8 is Science and Technology, with a success rate of approximately 79%.

| Task Name | <i>p</i> |
|------------------------------|----------|
| Concepts of Physical Science | 0.69 |
| Concepts of Life Science | 0.64 |
| Concepts of Earth Science | 0.65 |
| Science and Technology | 0.79 |

Science Tasks Grade 10

The most difficult task in Grade 10 is Concepts of Earth Science, with a success rate of approximately 65%. The easiest task in Grade 10 is Science and Technology, with a success rate of approximately 84%.

| Task Name | <i>p</i> |
|------------------------------|----------|
| Concepts of Physical Science | 0.71 |
| Concepts of Life Science | 0.77 |
| Concepts of Earth Science | 0.65 |
| Science and Technology | 0.84 |

Item Difficulty Analyses (Standard, No ELOS)

All item difficulty statistics for reading, writing, mathematics, and science are presented in the tables below. Item labels are provided, as are p -values. The p -value represents the proportion of the students responding in the keyed direction (e.g., students who received partial or full credit, with students receiving full credit contributing more significantly to the rating). The variable ' p ' is an index of item difficulty, with lower numbers indicating lower rates of success.

Reading Item Difficulties

The following tables summarize item difficulties for each grade band in reading. Reading item difficulties range from .30 to .90.

Reading Items Grade Band 3/4

The most difficult item in the 3/4 Grade Band is 1.34A, item 8, with a success rate of approximately 30%. The easiest item in the 3/4 Grade Band is 1.34A, item 6, with a success rate of approximately 85%.

| Item Description | p |
|--|------|
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_1 | 0.79 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_2 | 0.52 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_3 | 0.51 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_4 | 0.45 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_5 | 0.74 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_6 | 0.85 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_7 | 0.45 |
| Reading_Grade_3/4_Task_1.34A_Identify_Signs_and_Symbols_Item_8 | 0.30 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_1 | 0.68 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_2 | 0.49 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_3 | 0.56 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_4 | 0.74 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_5 | 0.81 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_6 | 0.67 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_7 | 0.74 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_8 | 0.72 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_9 | 0.62 |
| Reading_Grade_3/4_Task_1.34B_Identify_Letter_Sounds_Item_10 | 0.65 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_1 | 0.50 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_2 | 0.58 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_3 | 0.53 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_4 | 0.66 |

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_5 | 0.64 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_6 | 0.56 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_7 | 0.61 |
| Reading_Grade_3/4_Task_1.34C_Blend_Sounds_Item_8 | 0.59 |
| Reading_Grade_3/4_Task_1.34D_Identify_Own_Name_Item_1 | 0.87 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_1 | 0.57 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_2 | 0.63 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_3 | 0.45 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_4 | 0.38 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_5 | 0.69 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_6 | 0.66 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_7 | 0.70 |
| Reading_Grade_3/4_Task_2.34A_Read_Passages:_Story_1,_Annie_Goes_to_a_Party_Item_8 | 0.59 |
| Reading_Grade_3/4_Task_2.34B_Read_Passages:_Story_2,_Jill_and_the_Zoo_Item_1 | 0.56 |
| Reading_Grade_3/4_Task_2.34B_Read_Passages:_Story_2,_Jill_and_the_Zoo_Item_2 | 0.53 |
| Reading_Grade_3/4_Task_2.34B_Read_Passages:_Story_2,_Jill_and_the_Zoo_Item_3 | 0.69 |
| Reading_Grade_3/4_Task_2.34B_Read_Passages:_Story_2,_Jill_and_the_Zoo_Item_4 | 0.58 |

Reading Items Grade Band 5/6

The most difficult item in the 5/6 Grade Band is 1.56A, item 6, with a success rate of approximately 44%. The easiest item in the 5/6 Grade Band is 2.56B, item 1, with a success rate of approximately 90%.

| Item Description | <i>p</i> |
|--|----------|
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_1 | 0.67 |
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_2 | 0.48 |

| Item Description | <i>p</i> |
|--|----------|
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_3 | 0.54 |
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_4 | 0.62 |
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_5 | 0.78 |
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_6 | 0.82 |
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_7 | 0.68 |
| Reading_Grade_5/6_Task_1.56A_Read_Words_Item_8 | 0.44 |
| Reading_Grade_5/6_Task_1.56B_Read_Sentences_Item_1 | 0.53 |
| Reading_Grade_5/6_Task_1.56B_Read_Sentences_Item_2 | 0.49 |
| Reading_Grade_5/6_Task_1.56B_Read_Sentences_Item_3 | 0.51 |
| Reading_Grade_5/6_Task_1.56B_Read_Sentences_Item_4 | 0.58 |
| Reading_Grade_5/6_Task_1.56B_Read_Sentences_Item_5 | 0.46 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_1 | 0.83 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_2 | 0.59 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_3 | 0.59 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_4 | 0.73 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_5 | 0.58 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_6 | 0.68 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_7 | 0.65 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_8 | 0.58 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_9 | 0.50 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_10 | 0.61 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_11 | 0.87 |
| Reading_Grade_5/6_Task_2.56A_Read_Passages:_Story_1,_Jimmy_Rides_the_Bus_Item_12 | 0.46 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Item_1 | 0.90 |

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_2 | 0.68 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_3 | 0.65 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_4 | 0.76 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_5 | 0.71 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_6 | 0.68 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_7 | 0.64 |
| Reading_Grade_5/6_Task_2.56B_Read_Passages:_Story_2,_Jack_is_so_Friendly_Ite m_8 | 0.64 |
| Reading_Grade_5/6_Task_2.56C_Story_2,_Jack_is_so_Friendly_Item_1 | 0.50 |
| Reading_Grade_5/6_Task_2.56C_Story_2,_Jack_is_so_Friendly_Item_2 | 0.50 |

Grade Band 7/8

The most difficult item in the 7/8 Grade Band is 1.78B, item 1, with a success rate of approximately 40%. The easiest item in the 7/8 Grade Band is 2.78B, item 5, with a success rate of approximately 80%.

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_7/8_Task_1.78A_Read_Words_of_Increasing_Complexity_Item_1 | 0.61 |
| Reading_Grade_7/8_Task_1.78A_Read_Words_of_Increasing_Complexity_Item_2 | 0.59 |
| Reading_Grade_7/8_Task_1.78A_Read_Words_of_Increasing_Complexity_Item_3 | 0.60 |
| Reading_Grade_7/8_Task_1.78A_Read_Words_of_Increasing_Complexity_Item_4 | 0.74 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_1 | 0.40 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_2 | 0.52 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_3 | 0.43 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_4 | 0.50 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_5 | 0.76 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_6 | 0.73 |
| Reading_Grade_7/8_Task_1.78B_Obtain_Information_Item_7 | 0.55 |
| Reading_Grade_7/8_Task_1.78C_Read_Sentences_Item_1 | 0.58 |
| Reading_Grade_7/8_Task_1.78C_Read_Sentences_Item_2 | 0.60 |
| Reading_Grade_7/8_Task_1.78C_Read_Sentences_Item_3 | 0.55 |
| Reading_Grade_7/8_Task_1.78C_Read_Sentences_Item_4 | 0.66 |

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_1 | 0.78 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_2 | 0.70 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_3 | 0.67 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_4 | 0.60 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_5 | 0.68 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_6 | 0.58 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_7 | 0.68 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_8 | 0.65 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_9 | 0.66 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_10 | 0.79 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_11 | 0.74 |
| Reading_Grade_7/8_Task_2.78A_Read_Passages:_Story_1,_Hannah's_Homework_Item_12 | 0.49 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_1 | 0.78 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_2 | 0.69 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_3 | 0.71 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_4 | 0.79 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_5 | 0.80 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_6 | 0.74 |
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_7 | 0.76 |

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_7/8_Task_2.78B_Read_Passages:_Story_2,_Eating_Lunch_at_the_Cafeteria_Item_8 | 0.73 |

Reading Items Grade Band 9/10

The most difficult item in the 9/10 Grade Band is 2.910B, item 10, with a success rate of approximately 44%. The easiest item in the 9/10 Grade Band is 2.910B, item 8, with a success rate of approximately 83%.

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_1 | 0.62 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_2 | 0.65 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_3 | 0.62 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_4 | 0.77 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_5 | 0.72 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_6 | 0.70 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_7 | 0.74 |
| Reading_Grade_9/10_Task_1.910A_Decode_Words_Item_8 | 0.72 |
| Reading_Grade_9/10_Task_1.910B_Identify_Root_Words_Item_1 | 0.64 |
| Reading_Grade_9/10_Task_1.910B_Identify_Root_Words_Item_2 | 0.69 |
| Reading_Grade_9/10_Task_1.910B_Identify_Root_Words_Item_3 | 0.61 |
| Reading_Grade_9/10_Task_1.910B_Identify_Root_Words_Item_4 | 0.55 |
| Reading_Grade_9/10_Task_1.910B_Identify_Root_Words_Item_5 | 0.59 |
| Reading_Grade_9/10_Task_1.910B_Identify_Root_Words_Item_6 | 0.66 |
| Reading_Grade_9/10_Task_1.910C_Follow_Multi-Step_Directions_Item_1 | 0.66 |
| Reading_Grade_9/10_Task_1.910C_Follow_Multi-Step_Directions_Item_2 | 0.64 |
| Reading_Grade_9/10_Task_1.910C_Follow_Multi-Step_Directions_Item_3 | 0.66 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_Item_1 | 0.46 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_Item_2 | 0.76 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_Item_3 | 0.69 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_Item_4 | 0.68 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_Item_5 | 0.65 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_Item_6 | 0.58 |

| Item Description | <i>p</i> |
|---|----------|
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_It em_7 | 0.57 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_It em_8 | 0.64 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_It em_9 | 0.49 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_It em_10 | 0.50 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_It em_11 | 0.60 |
| Reading_Grade_9/10_Task_2.910A_Read_Passages:_Story_1,_Jan_and_the_Party_It em_12 | 0.45 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_1 | 0.69 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_2 | 0.68 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_3 | 0.81 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_4 | 0.71 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_5 | 0.74 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_6 | 0.80 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_7 | 0.68 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_8 | 0.83 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_9 | 0.56 |
| Reading_Grade_9/10_Task_2.910B_Read_Passages:_Story_2,_Ready_for_Graduatio n_Item_10 | 0.44 |
| Reading_Grade_9/10_Task_2.910C_Story_2,_Ready_for_Graduation_Item_1 | 0.55 |
| Reading_Grade_9/10_Task_2.910C_Story_2,_Ready_for_Graduation_Item_2 | 0.46 |

Writing Item Difficulties

The following tables summarize item difficulties for each grade band in writing. Writing item difficulties range from .35 to .84.

Writing Items Grade Band 3/4

The most difficult item in the 3/4 Grade Band is Copy Words, item 3, with a success rate of approximately 51%. The easiest item in the 3/4 Grade Band is Copy Letters, Item 10, with a success rate of approximately 77%.

| Item Description | <i>p</i> |
|--|----------|
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_1 | 0.59 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_2 | 0.59 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_3 | 0.59 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_4 | 0.76 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_5 | 0.71 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_6 | 0.68 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_7 | 0.63 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_8 | 0.77 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_9 | 0.72 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_10 | 0.77 |
| Writing_Grade_3/4_Task_1.34A_Copy_Letters_Item_11 | 0.65 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_1 | 0.55 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_2 | 0.55 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_3 | 0.51 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_4 | 0.69 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_5 | 0.65 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_6 | 0.7 |
| Writing_Grade_3/4_Task_1.34B_Copy_Words_Item_7 | 0.74 |
| Writing_Grade_3/4_Task_1.34C_Write_Own_Name_Item_1 | 0.59 |
| Writing_Grade_3/4_Task_1.34D_Matching_and_Sequencing_Pictures_Item_1 | 0.65 |
| Writing_Grade_3/4_Task_1.34D_Matching_and_Sequencing_Pictures_Item_2 | 0.57 |

Writing Items Grade Band 5/6

The most difficult item in the 5/6 Grade Band is Write Words from Dictation, Item 5, with a success rate of approximately 42%. The easiest item in the 5/6 Grade Band is Write Words from Dictation, Item 4, with a success rate of approximately 66%.

| Item Description | <i>p</i> |
|--|----------|
| Writing_Grade_5/6_Task_1.56A_Conventions_of_Writing_Item_1 | 0.62 |
| Writing_Grade_5/6_Task_1.56A_Conventions_of_Writing_Item_2 | 0.51 |
| Writing_Grade_5/6_Task_1.56A_Conventions_of_Writing_Item_3 | 0.64 |
| Writing_Grade_5/6_Task_1.56A_Conventions_of_Writing_Item_4 | 0.48 |
| Writing_Grade_5/6_Task_1.56B_Write_Own_Name_Item_1 | 0.65 |
| Writing_Grade_5/6_Task_1.56C_Write_Words_from_Dictation_Item_1 | 0.56 |
| Writing_Grade_5/6_Task_1.56C_Write_Words_from_Dictation_Item_2 | 0.60 |
| Writing_Grade_5/6_Task_1.56C_Write_Words_from_Dictation_Item_3 | 0.52 |
| Writing_Grade_5/6_Task_1.56C_Write_Words_from_Dictation_Item_4 | 0.66 |
| Writing_Grade_5/6_Task_1.56C_Write_Words_from_Dictation_Item_5 | 0.42 |
| Writing_Grade_5/6_Task_1.56D_Write_a_Sentence_Item_1 | 0.62 |
| Writing_Grade_5/6_Task_1.56D_Write_a_Sentence_Item_2 | 0.63 |
| Writing_Grade_5/6_Task_1.56D_Write_a_Sentence_Item_3 | 0.63 |

Writing Grade Band 7/8

The most difficult item in the 7/8 Grade Band is Write a Sentence, Item 2, with a success rate of approximately 35%. The easiest item in the 7/8 Grade Band is Revise Sentences, Item 4, with a success rate of approximately 84%.

| Item Description | <i>p</i> |
|---|----------|
| Writing_Grade_7/8_Task_1.78A_Write_Sentences_from_Dictation_Item_1 | 0.43 |
| Writing_Grade_7/8_Task_1.78A_Write_Sentences_from_Dictation_Item_2 | 0.41 |
| Writing_Grade_7/8_Task_1.78A_Write_Sentences_from_Dictation_Item_3 | 0.44 |
| Writing_Grade_7/8_Task_1.78B_Conventions_of_Standard_English_Item_1 | 0.58 |
| Writing_Grade_7/8_Task_1.78B_Conventions_of_Standard_English_Item_2 | 0.43 |
| Writing_Grade_7/8_Task_1.78B_Conventions_of_Standard_English_Item_3 | 0.46 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_1 | 0.44 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_2 | 0.57 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_3 | 0.46 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_4 | 0.65 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_5 | 0.63 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_6 | 0.62 |
| Writing_Grade_7/8_Task_1.78C_Communicate_Ideas_Using_Words_Item_7 | 0.64 |
| Writing_Grade_7/8_Task_1.78D_Write_a_Sentence_Item_1 | 0.38 |
| Writing_Grade_7/8_Task_1.78D_Write_a_Sentence_Item_2 | 0.35 |
| Writing_Grade_7/8_Task_1.78E_Revise_Sentences_Item_1 | 0.69 |
| Writing_Grade_7/8_Task_1.78E_Revise_Sentences_Item_2 | 0.72 |
| Writing_Grade_7/8_Task_1.78E_Revise_Sentences_Item_3 | 0.67 |
| Writing_Grade_7/8_Task_1.78E_Revise_Sentences_Item_4 | 0.84 |

Writing Items Grade Band 9/10

The most difficult item in the 9/10 Grade Band is Write a Story, Item 1, with a success rate of approximately 40%. The easiest item in the 9/10 Grade Band is Revise Writing, Item 9, with a success rate of approximately 79%.

| Item Description | <i>p</i> |
|---|----------|
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_1 | 0.73 |
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_2 | 0.59 |
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_3 | 0.61 |
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_4 | 0.77 |
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_5 | 0.77 |
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_6 | 0.76 |
| Writing_Grade_9/10_Task_1.910A_Conventions_of_Standard_English_Item_7 | 0.65 |
| Writing_Grade_9/10_Task_1.910B_Write_a_Story_Item_1 | 0.40 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_1 | 0.71 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_2 | 0.68 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_3 | 0.63 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_4 | 0.75 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_5 | 0.72 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_6 | 0.73 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_7 | 0.63 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_8 | 0.76 |
| Writing_Grade_9/10_Task_1.910C_Revise_Writing_Item_9 | 0.79 |

Mathematics Item Difficulties

The following tables summarize item difficulties for each grade band in mathematics. Math item difficulties range from .21 to .99.

Mathematics Items Grade Band 3/4

The most difficult item in the 3/4 Grade Band is 1.34B, Item 2, with a success rate of approximately 36%. The easiest items in the 3/4 Grade Band are 1.34A, Items 6 & 7, with a success rate of approximately 81%.

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_1 | 0.58 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_2 | 0.62 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_3 | 0.60 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_4 | 0.73 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_5 | 0.72 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_6 | 0.81 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_7 | 0.81 |
| Math_Grade_3/4_Task_1.34A_Copy_Numbers_Item_8 | 0.70 |
| Math_Grade_3/4_Task_1.34B_First_and_Last_Item_1 | 0.64 |
| Math_Grade_3/4_Task_1.34B_First_and_Last_Item_2 | 0.36 |
| Math_Grade_3/4_Task_1.34B_First_and_Last_Item_3 | 0.58 |
| Math_Grade_3/4_Task_1.34B_First_and_Last_Item_4 | 0.63 |
| Math_Grade_3/4_Task_1.34B_First_and_Last_Item_5 | 0.50 |
| Math_Grade_3/4_Task_1.34B_First_and_Last_Item_6 | 0.51 |
| Math_Grade_3/4_Task_1.34C_Count_Item_1 | 0.70 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_1 | 0.63 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_2 | 0.59 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_3 | 0.68 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_4 | 0.68 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_5 | 0.78 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_6 | 0.74 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_7 | 0.79 |
| Math_Grade_3/4_Task_2.34_Same/_Different_Item_8 | 0.67 |
| Math_Grade_3/4_Task_3.34_Identify_Shapes_Item_1 | 0.72 |
| Math_Grade_3/4_Task_3.34_Identify_Shapes_Item_2 | 0.78 |
| Math_Grade_3/4_Task_3.34_Identify_Shapes_Item_3 | 0.70 |

Mathematics Items Grade Band 5/6

The most difficult item in the 5/6 Grade Band is 6.56C, item 1, with a success rate of approximately 36%. The easiest item in the 5/6 Grade Band is 6.56A, item 1, with a success rate of approximately 92%.

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_5/6_Task_1.56A_Read_and_Write_Numbers_Item_1 | 0.80 |
| Math_Grade_5/6_Task_1.56A_Read_and_Write_Numbers_Item_2 | 0.79 |
| Math_Grade_5/6_Task_1.56A_Read_and_Write_Numbers_Item_3 | 0.71 |
| Math_Grade_5/6_Task_1.56A_Read_and_Write_Numbers_Item_4 | 0.88 |
| Math_Grade_5/6_Task_1.56A_Read_and_Write_Numbers_Item_5 | 0.82 |
| Math_Grade_5/6_Task_1.56B_Number_Line,First_and_Last_Item_1 | 0.68 |
| Math_Grade_5/6_Task_1.56B_Number_Line,First_and_Last_Item_2 | 0.61 |
| Math_Grade_5/6_Task_1.56B_Number_Line,First_and_Last_Item_3 | 0.51 |
| Math_Grade_5/6_Task_1.56B_Number_Line,First_and_Last_Item_4 | 0.80 |
| Math_Grade_5/6_Task_1.56B_Number_Line,First_and_Last_Item_5 | 0.63 |
| Math_Grade_5/6_Task_1.56B_Number_Line,First_and_Last_Item_6 | 0.64 |
| Math_Grade_5/6_Task_1.56C_Count_Objects_Item_1 | 0.73 |
| Math_Grade_5/6_Task_1.56D_Count_Item_1 | 0.78 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_1 | 0.60 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_2 | 0.60 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_3 | 0.52 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_4 | 0.69 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_5 | 0.75 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_6 | 0.69 |
| Math_Grade_5/6_Task_2.56_Simple_Addition_Item_7 | 0.57 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_1 | 0.76 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_2 | 0.73 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_3 | 0.79 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_4 | 0.83 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_5 | 0.82 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_6 | 0.83 |
| Math_Grade_5/6_Task_3.56_Reproduce_Simple_Patterns_Item_7 | 0.57 |
| Math_Grade_5/6_Task_4.56_Read_Simple_Graphs_Item_1 | 0.46 |
| Math_Grade_5/6_Task_4.56_Read_Simple_Graphs_Item_2 | 0.38 |
| Math_Grade_5/6_Task_4.56_Read_Simple_Graphs_Item_3 | 0.73 |
| Math_Grade_5/6_Task_4.56_Read_Simple_Graphs_Item_4 | 0.84 |
| Math_Grade_5/6_Task_4.56_Read_Simple_Graphs_Item_5 | 0.84 |
| Math_Grade_5/6_Task_4.56_Read_Simple_Graphs_Item_6 | 0.87 |
| Math_Grade_5/6_Task_5.56A_Same,_Longer/Shorter,_More_Item_1 | 0.74 |

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_5/6_Task_5.56A_Same,_Longer/Shorter,_More_Item_2 | 0.82 |
| Math_Grade_5/6_Task_5.56A_Same,_Longer/Shorter,_More_Item_3 | 0.86 |
| Math_Grade_5/6_Task_5.56B_Identify_Money_Item_1 | 0.86 |
| Math_Grade_5/6_Task_5.56B_Identify_Money_Item_2 | 0.59 |
| Math_Grade_5/6_Task_5.56B_Identify_Money_Item_3 | 0.72 |
| Math_Grade_5/6_Task_5.56B_Identify_Money_Item_4 | 0.75 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_1 | 0.92 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_2 | 0.87 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_3 | 0.84 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_4 | 0.87 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_5 | 0.83 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_6 | 0.86 |
| Math_Grade_5/6_Task_6.56A_Identify_Shapes_Item_7 | 0.75 |
| Math_Grade_5/6_Task_6.56B_Same_or_Different_(Shapes)_Item_1 | 0.74 |
| Math_Grade_5/6_Task_6.56B_Same_or_Different_(Shapes)_Item_2 | 0.89 |
| Math_Grade_5/6_Task_6.56B_Same_or_Different_(Shapes)_Item_3 | 0.73 |
| Math_Grade_5/6_Task_6.56C_Identify_Perimeter_Item_1 | 0.36 |

Mathematics Grade Band 7/8

The most difficult item in the 7/8 Grade Band is 6.78B, item 3, with a success rate of approximately 21%. The easiest item in the 7/8 Grade Band is 4.78, item 5, with a success rate of approximately 94%.

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_7/8_Task_1.78A_Read_and_Write_Numbers,_ID_Place_Value_Item_1 | 0.65 |
| Math_Grade_7/8_Task_1.78A_Read_and_Write_Numbers,_ID_Place_Value_Item_2 | 0.72 |
| Math_Grade_7/8_Task_1.78A_Read_and_Write_Numbers,_ID_Place_Value_Item_3 | 0.37 |
| Math_Grade_7/8_Task_1.78A_Read_and_Write_Numbers,_ID_Place_Value_Item_4 | 0.56 |
| Math_Grade_7/8_Task_1.78B_Identify_Fractions_Item_1 | 0.65 |
| Math_Grade_7/8_Task_1.78B_Identify_Fractions_Item_2 | 0.48 |
| Math_Grade_7/8_Task_1.78C_Ordering_-_Number_Line_and_Pictures_Item_1 | 0.70 |
| Math_Grade_7/8_Task_1.78C_Ordering_-_Number_Line_and_Pictures_Item_2 | 0.64 |
| Math_Grade_7/8_Task_1.78C_Ordering_-_Number_Line_and_Pictures_Item_3 | 0.58 |
| Math_Grade_7/8_Task_1.78C_Ordering_-_Number_Line_and_Pictures_Item_4 | 0.53 |
| Math_Grade_7/8_Task_1.78C_Ordering_-_Number_Line_and_Pictures_Item_5 | 0.56 |
| Math_Grade_7/8_Task_1.78D_Identify_SkiPatterns_Item_1 | 0.60 |
| Math_Grade_7/8_Task_1.78E_Count_Item_1 | 0.73 |
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_1 | 0.55 |
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_2 | 0.58 |

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_3 | 0.54 |
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_4 | 0.70 |
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_5 | 0.60 |
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_6 | 0.59 |
| Math_Grade_7/8_Task_2.78_Double_Digit_Addition_and_Subtraction_Item_7 | 0.50 |
| Math_Grade_7/8_Task_3.78A_Reproduce_and_Extend_Simple_Patterns_Item_1 | 0.75 |
| Math_Grade_7/8_Task_3.78A_Reproduce_and_Extend_Simple_Patterns_Item_2 | 0.72 |
| Math_Grade_7/8_Task_3.78A_Reproduce_and_Extend_Simple_Patterns_Item_3 | 0.63 |
| Math_Grade_7/8_Task_3.78B_Label_a_Set_as_None_or_Zero_Item_1 | 0.73 |
| Math_Grade_7/8_Task_3.78B_Label_a_Set_as_None_or_Zero_Item_2 | 0.74 |
| Math_Grade_7/8_Task_3.78B_Label_a_Set_as_None_or_Zero_Item_3 | 0.74 |
| Math_Grade_7/8_Task_3.78B_Label_a_Set_as_None_or_Zero_Item_4 | 0.82 |
| Math_Grade_7/8_Task_3.78C_Understand_Symbols_Item_1 | 0.86 |
| Math_Grade_7/8_Task_3.78C_Understand_Symbols_Item_2 | 0.80 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_1 | 0.51 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_2 | 0.70 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_3 | 0.86 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_4 | 0.83 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_5 | 0.94 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_6 | 0.89 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_7 | 0.62 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_8 | 0.58 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_9 | 0.66 |
| Math_Grade_7/8_Task_4.78_Read_Simple_Graphs_Item_10 | 0.54 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_1 | 0.41 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_2 | 0.64 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_3 | 0.74 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_4 | 0.63 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_5 | 0.78 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_6 | 0.71 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_7 | 0.74 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_8 | 0.65 |
| Math_Grade_7/8_Task_5.78A_Identify_Units_of_Measurement_Item_9 | 0.55 |
| Math_Grade_7/8_Task_5.78B_Count_Money_Item_1 | 0.42 |
| Math_Grade_7/8_Task_5.78B_Count_Money_Item_2 | 0.91 |
| Math_Grade_7/8_Task_5.78C_Identify_Money_Item_1 | 0.87 |
| Math_Grade_7/8_Task_5.78C_Identify_Money_Item_2 | 0.88 |
| Math_Grade_7/8_Task_5.78C_Identify_Money_Item_3 | 0.87 |
| Math_Grade_7/8_Task_5.78C_Identify_Money_Item_4 | 0.94 |

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_1 | 0.89 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_2 | 0.79 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_3 | 0.86 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_4 | 0.86 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_5 | 0.86 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_6 | 0.77 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_7 | 0.76 |
| Math_Grade_7/8_Task_6.78A_Identify_Shapes_and_Position_Item_8 | 0.93 |
| Math_Grade_7/8_Task_6.78B_Match_Shapes_Item_1 | 0.93 |
| Math_Grade_7/8_Task_6.78B_Match_Shapes_Item_2 | 0.93 |
| Math_Grade_7/8_Task_6.78B_Match_Shapes_Item_3 | 0.21 |
| Math_Grade_7/8_Task_6.78C_Identify_Perimeter_Item_1 | 0.57 |

Mathematics Items Grade Band 9/10

The most difficult item in the 9/10 Grade Band is 6.910A, item 11, with a success rate of approximately 33%. The easiest item in the 9/10 Grade Band is 6.910A, item 5, with a success rate of approximately 99%.

| Item Description | <i>p</i> |
|--|----------|
| Math_Grade_9/10_Task_1.910A_Identify_Place_Values_Item_1 | 0.60 |
| Math_Grade_9/10_Task_1.910A_Identify_Place_Values_Item_2 | 0.65 |
| Math_Grade_9/10_Task_1.910A_Identify_Place_Values_Item_3 | 0.66 |
| Math_Grade_9/10_Task_1.910A_Identify_Place_Values_Item_4 | 0.75 |
| Math_Grade_9/10_Task_1.910A_Identify_Place_Values_Item_5 | 0.79 |
| Math_Grade_9/10_Task_1.910B_Identify_Fractions_Item_1 | 0.70 |
| Math_Grade_9/10_Task_1.910B_Identify_Fractions_Item_2 | 0.52 |
| Math_Grade_9/10_Task_1.910B_Identify_Fractions_Item_3 | 0.55 |
| Math_Grade_9/10_Task_1.910B_Identify_Fractions_Item_4 | 0.75 |
| Math_Grade_9/10_Task_1.910C_Order_Numbers_Item_1 | 0.45 |
| Math_Grade_9/10_Task_2.910A_Round_Numbers_Item_1 | 0.53 |
| Math_Grade_9/10_Task_2.910A_Round_Numbers_Item_2 | 0.43 |
| Math_Grade_9/10_Task_2.910A_Round_Numbers_Item_3 | 0.59 |
| Math_Grade_9/10_Task_2.910B_Addition/Subtraction,_Multiplication_Item_1 | 0.68 |
| Math_Grade_9/10_Task_2.910B_Addition/Subtraction,_Multiplication_Item_2 | 0.70 |
| Math_Grade_9/10_Task_2.910B_Addition/Subtraction,_Multiplication_Item_3 | 0.71 |
| Math_Grade_9/10_Task_2.910B_Addition/Subtraction,_Multiplication_Item_4 | 0.62 |
| Math_Grade_9/10_Task_2.910B_Addition/Subtraction,_Multiplication_Item_5 | 0.57 |
| Math_Grade_9/10_Task_2.910B_Addition/Subtraction,_Multiplication_Item_6 | 0.86 |
| Math_Grade_9/10_Task_3.910A_Extend_a_Pattern/Supply_Missing_Element_Item_1 | 0.55 |
| Math_Grade_9/10_Task_3.910A_Extend_a_Pattern/Supply_Missing_Element_Item_2 | 0.63 |
| Math_Grade_9/10_Task_3.910A_Extend_a_Pattern/Supply_Missing_Element_Item_3 | 0.86 |
| Math_Grade_9/10_Task_3.910A_Extend_a_Pattern/Supply_Missing_Element_Item_4 | 0.91 |
| Math_Grade_9/10_Task_3.910A_Extend_a_Pattern/Supply_Missing_Element_Item_5 | 0.64 |
| Math_Grade_9/10_Task_3.910A_Extend_a_Pattern/Supply_Missing_Element_Item_6 | 0.64 |
| Math_Grade_9/10_Task_3.910B_Understand_Symbols_Item_1 | 0.70 |
| Math_Grade_9/10_Task_3.910B_Understand_Symbols_Item_2 | 0.51 |
| Math_Grade_9/10_Task_3.910B_Understand_Symbols_Item_3 | 0.54 |
| Math_Grade_9/10_Task_3.910B_Understand_Symbols_Item_4 | 0.60 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_1 | 0.85 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_2 | 0.93 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_3 | 0.89 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_4 | 0.95 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_5 | 0.90 |

| Item Description | <i>p</i> |
|---|----------|
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_6 | 0.74 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_7 | 0.73 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_8 | 0.79 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_9 | 0.73 |
| Math_Grade_9/10_Task_4.910_Read_Simple_Graphs_Item_10 | 0.62 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_1 | 0.69 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_2 | 0.89 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_3 | 0.81 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_4 | 0.78 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_5 | 0.71 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_6 | 0.81 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_7 | 0.74 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_8 | 0.57 |
| Math_Grade_9/10_Task_5.910A_Identify_Units_of_Measurement_Item_9 | 0.71 |
| Math_Grade_9/10_Task_5.910B_Count_Money_Item_1 | 0.59 |
| Math_Grade_9/10_Task_5.910B_Count_Money_Item_2 | 0.91 |
| Math_Grade_9/10_Task_5.910B_Count_Money_Item_3 | 0.70 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_1 | 0.90 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_2 | 0.88 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_3 | 0.80 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_4 | 0.79 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_5 | 0.99 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_6 | 0.80 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_7 | 0.86 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_8 | 0.92 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_9 | 0.67 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_10 | 0.56 |
| Math_Grade_9/10_Task_6.910A_Describe_and_Compare_Shapes,_Shapes_Greater,_Less,_Equal_To_Item_11 | 0.33 |

Science Item Difficulties

The following tables summarize item difficulties for each grade band in science. Science item difficulties range from .40 to .95.

Science Items Grade 4

The most difficult item in Grade 4 is 1.4, Item 4, with a success rate of approximately 40%. The easiest item in Grade 4 is 1.4, Item 5, with a success rate of approximately 88%.

| Item Description | <i>p</i> |
|---|----------|
| Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_1 | 0.71 |
| Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_2 | 0.65 |
| Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_3 | 0.78 |
| Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_4 | 0.40 |
| Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_5 | 0.88 |
| Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_6 | 0.82 |
| Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_1 | 0.77 |
| Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_2 | 0.64 |
| Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_3 | 0.59 |
| Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_4 | 0.68 |
| Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_5 | 0.67 |
| Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_6 | 0.48 |
| Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_1 | 0.67 |
| Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_2 | 0.79 |
| Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_3 | 0.67 |
| Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_4 | 0.76 |
| Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_5 | 0.84 |
| Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_6 | 0.81 |
| Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_a nd_Technology_Item_1 | 0.76 |
| Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_a nd_Technology_Item_2 | 0.67 |
| Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_a nd_Technology_Item_3 | 0.64 |
| Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_a nd_Technology_Item_4 | 0.86 |
| Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_a nd_Technology_Item_5 | 0.69 |
| Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_a nd_Technology_Item_6 | 0.87 |

Science Items Grade 8

The most difficult item in Grade 8 is 3.8, item 5, with a success rate of approximately 58%. The easiest item in Grade 8 is 3.8, item 6, with a success rate of approximately 91%.

| Item Description | <i>p</i> |
|--|----------|
| Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_1 | 0.80 |
| Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_2 | 0.79 |
| Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_3 | 0.64 |
| Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_4 | 0.72 |
| Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_5 | 0.70 |
| Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_6 | 0.84 |
| Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_1 | 0.69 |
| Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_2 | 0.64 |
| Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_3 | 0.70 |
| Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_4 | 0.61 |
| Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_5 | 0.74 |
| Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_6 | 0.83 |
| Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_1 | 0.70 |
| Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_2 | 0.69 |
| Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_3 | 0.73 |
| Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_4 | 0.70 |
| Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_5 | 0.58 |
| Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_6 | 0.91 |
| Science_Grade_8_Task_4.8_Science_and_Technology_Item_1 | 0.85 |
| Science_Grade_8_Task_4.8_Science_and_Technology_Item_2 | 0.88 |
| Science_Grade_8_Task_4.8_Science_and_Technology_Item_3 | 0.75 |
| Science_Grade_8_Task_4.8_Science_and_Technology_Item_4 | 0.80 |
| Science_Grade_8_Task_4.8_Science_and_Technology_Item_5 | 0.84 |
| Science_Grade_8_Task_4.8_Science_and_Technology_Item_6 | 0.72 |

Science Items Grade 10

The most difficult items in Grade 10 are 1.10, item 6 and 3.10, item 6, with success rates of approximately 45%. The easiest item in Grade 10 is 4.10, item 6, with a success rate of approximately 95%.

| Item Description | <i>p</i> |
|--|----------|
| Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_1 | 0.77 |
| Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_2 | 0.79 |
| Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_3 | 0.69 |
| Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_4 | 0.89 |
| Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_5 | 0.68 |
| Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_6 | 0.45 |
| Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_1 | 0.86 |
| Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_2 | 0.65 |
| Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_3 | 0.88 |
| Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_4 | 0.55 |
| Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_5 | 0.89 |
| Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_6 | 0.86 |
| Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_1 | 0.74 |
| Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_2 | 0.54 |
| Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_3 | 0.76 |
| Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_4 | 0.54 |
| Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_5 | 0.89 |
| Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_6 | 0.45 |
| Science_Grade_10_Task_4.10_Science_and_Technology_Item_1 | 0.69 |
| Science_Grade_10_Task_4.10_Science_and_Technology_Item_2 | 0.95 |
| Science_Grade_10_Task_4.10_Science_and_Technology_Item_3 | 0.77 |
| Science_Grade_10_Task_4.10_Science_and_Technology_Item_4 | 0.77 |
| Science_Grade_10_Task_4.10_Science_and_Technology_Item_5 | 0.93 |
| Science_Grade_10_Task_4.10_Science_and_Technology_Item_6 | 0.95 |

ELOS Task Level of Support Analysis

Each task in reading, writing, and mathematics was analyzed across the 3-4, 5-6, 7-8, and 9-10 grade bands to determine the level of support required to achieve success, on average. Level 1 is full physical support; level 2 is partial physical contact support; level 3 is visual, verbal and/or gestural support; level 4 is complete independence. Science was analyzed in the same manner for grades 4, 8, and 10. Task 2 results are bolded to allow for visual discrimination of the results. The results demonstrate a clear pattern that most students who are participating on the ELOS assessments are functioning in two separate groups, those needing full physical support (Level 1) and those not needing any support (Level 4)

to complete the assigned tasks. Fewer students across the gamut of results required a Level 2 support system (partial physical support) or a Level 3 support system (visual/ verbal/ gestural support). Please note that the results provided are rounded and thus each task may not add up to exactly 100%.

ELOS Reading Grades 3-4

| Task Number | Level of Support (Loss)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|---|---|
| Task 1 | 1 | 38% |
| Task 1 | 2 | 11% |
| Task 1 | 3 | 12% |
| Task 1 | 4 | 39% |
| Task 2 | 1 | 39% |
| Task 2 | 2 | 12% |
| Task 2 | 3 | 18% |
| Task 2 | 4 | 31% |
| Task 3 | 1 | 35% |
| Task 3 | 2 | 10% |
| Task 3 | 3 | 14% |
| Task 3 | 4 | 41% |

ELOS Reading Grades 5-6

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 45% |
| Task 1 | 2 | 6% |
| Task 1 | 3 | 11% |
| Task 1 | 4 | 38% |
| Task 2 | 1 | 53% |
| Task 2 | 2 | 13% |
| Task 2 | 3 | 14% |
| Task 2 | 4 | 20% |
| Task 3 | 1 | 45% |
| Task 3 | 2 | 11% |
| Task 3 | 3 | 15% |
| Task 3 | 4 | 29% |

ELOS Reading Grades 7-8

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 31% |
| Task 1 | 2 | 7% |
| Task 1 | 3 | 15% |
| Task 1 | 4 | 47% |
| Task 2 | 1 | 43% |
| Task 2 | 2 | 7% |
| Task 2 | 3 | 21% |
| Task 2 | 4 | 29% |
| Task 3 | 1 | 37% |
| Task 3 | 2 | 12% |
| Task 3 | 3 | 17% |
| Task 3 | 4 | 33% |

ELOS Reading Grades 9-10

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 28% |
| Task 1 | 2 | 11% |
| Task 1 | 3 | 17% |
| Task 1 | 4 | 44% |
| Task 2 | 1 | 34% |
| Task 2 | 2 | 6% |
| Task 2 | 3 | 8% |
| Task 2 | 4 | 52% |
| Task 3 | 1 | 37% |
| Task 3 | 2 | 12% |
| Task 3 | 3 | 15% |
| Task 3 | 4 | 37% |

ELOS Writing Grades 3-4

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 41% |
| Task 1 | 2 | 10% |
| Task 1 | 3 | 12% |
| Task 1 | 4 | 37% |
| Task 2 | 1 | 46% |
| Task 2 | 2 | 12% |
| Task 2 | 3 | 10% |
| Task 2 | 4 | 33% |
| Task 3 | 1 | 37% |
| Task 3 | 2 | 12% |
| Task 3 | 3 | 9% |
| Task 3 | 4 | 43% |

ELOS Writing Grades 5-6

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 50% |
| Task 1 | 2 | 10% |
| Task 1 | 3 | 16% |
| Task 1 | 4 | 23% |
| Task 2 | 1 | 53% |
| Task 2 | 2 | 8% |
| Task 2 | 3 | 15% |
| Task 2 | 4 | 23% |
| Task 3 | 1 | 54% |
| Task 3 | 2 | 6% |
| Task 3 | 3 | 21% |
| Task 3 | 4 | 19% |

ELOS Writing Grades 7-8

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 33% |
| Task 1 | 2 | 9% |
| Task 1 | 3 | 14% |
| Task 1 | 4 | 44% |
| Task 2 | 1 | 56% |
| Task 2 | 2 | 6% |
| Task 2 | 3 | 11% |
| Task 2 | 4 | 27% |
| Task 3 | 1 | 52% |
| Task 3 | 2 | 6% |
| Task 3 | 3 | 16% |
| Task 3 | 4 | 26% |

ELOS Writing Grades 9-10

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 39% |
| Task 1 | 2 | 10% |
| Task 1 | 3 | 14% |
| Task 1 | 4 | 37% |
| Task 2 | 1 | 48% |
| Task 2 | 2 | 12% |
| Task 2 | 3 | 9% |
| Task 2 | 4 | 31% |
| Task 3 | 1 | 54% |
| Task 3 | 2 | 8% |
| Task 3 | 3 | 10% |
| Task 3 | 4 | 29% |

ELOS Mathematics Grades 3-4

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 44% |
| Task 1 | 2 | 8% |
| Task 1 | 3 | 12% |
| Task 1 | 4 | 36% |
| Task 2 | 1 | 47% |
| Task 2 | 2 | 5% |
| Task 2 | 3 | 15% |
| Task 2 | 4 | 32% |
| Task 3 | 1 | 40% |
| Task 3 | 2 | 5% |
| Task 3 | 3 | 16% |
| Task 3 | 4 | 38% |

ELOS Mathematics Grades 5-6

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 39% |
| Task 1 | 2 | 9% |
| Task 1 | 3 | 13% |
| Task 1 | 4 | 38% |
| Task 2 | 1 | 37% |
| Task 2 | 2 | 6% |
| Task 2 | 3 | 17% |
| Task 2 | 4 | 40% |
| Task 3 | 1 | 47% |
| Task 3 | 2 | 5% |
| Task 3 | 3 | 20% |
| Task 3 | 4 | 28% |

ELOS Mathematics Grades 7-8

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 35% |
| Task 1 | 2 | 7% |
| Task 1 | 3 | 10% |
| Task 1 | 4 | 47% |
| Task 2 | 1 | 37% |
| Task 2 | 2 | 5% |
| Task 2 | 3 | 13% |
| Task 2 | 4 | 44% |
| Task 3 | 1 | 41% |
| Task 3 | 2 | 7% |
| Task 3 | 3 | 14% |
| Task 3 | 4 | 38% |

ELOS Mathematics Grades 9-10

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 34% |
| Task 1 | 2 | 8% |
| Task 1 | 3 | 16% |
| Task 1 | 4 | 42% |
| Task 2 | 1 | 33% |
| Task 2 | 2 | 13% |
| Task 2 | 3 | 20% |
| Task 2 | 4 | 34% |
| Task 3 | 1 | 32% |
| Task 3 | 2 | 8% |
| Task 3 | 3 | 12% |
| Task 3 | 4 | 48% |

ELOS Science Grade 4

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 46% |
| Task 1 | 2 | 11% |
| Task 1 | 3 | 7% |
| Task 1 | 4 | 36% |
| Task 2 | 1 | 41% |
| Task 2 | 2 | 11% |
| Task 2 | 3 | 10% |
| Task 2 | 4 | 37% |
| Task 3 | 1 | 34% |
| Task 3 | 2 | 20% |
| Task 3 | 3 | 9% |
| Task 3 | 4 | 37% |

ELOS Science Grade 8

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 30% |
| Task 1 | 2 | 19% |
| Task 1 | 3 | 13% |
| Task 1 | 4 | 38% |
| Task 2 | 1 | 42% |
| Task 2 | 2 | 14% |
| Task 2 | 3 | 9% |
| Task 2 | 4 | 34% |
| Task 3 | 1 | 29% |
| Task 3 | 2 | 16% |
| Task 3 | 3 | 7% |
| Task 3 | 4 | 48% |

ELOS Science Grade 10

| Task Number | Level of Support (LoS)/ Range (1-4) | % of Students Who received this LoS on this Task |
|--------------------|--|---|
| Task 1 | 1 | 33% |
| Task 1 | 2 | 7% |
| Task 1 | 3 | 23% |
| Task 1 | 4 | 37% |
| Task 2 | 1 | 30% |
| Task 2 | 2 | 7% |
| Task 2 | 3 | 10% |
| Task 2 | 4 | 53% |
| Task 3 | 1 | 33% |
| Task 3 | 2 | 3% |
| Task 3 | 3 | 7% |
| Task 3 | 4 | 57% |

ELOS Item Order Analysis

The ELOS RWMS were also reviewed to determine whether the field test items developed for the 2011-12 assessments were functioning as intended. The items were designed such that within each of three tasks per grade band, per content area, item 1 should be less difficult than item 2, item 2 less difficult than item 3, and so forth. Item 1 was written as an attention item, item 2 as an interaction item, item 3 as an easy item, item 4 as a medium item, and item 5 as a difficult item.

Item difficulties were calculated using the average scores for all students on each item. A review of average item difficulties demonstrated that the test design worked remarkably well. Out of a total of 180 possible item sequences on all assessments, only 27 were deemed to be out of order in the domain of item difficulty, meaning that the item difficulties for subsequent items were either equal or higher. Of those 27 sequences, 4 had an exact match. This means that only 23 of the sequences had the items in the wrong order based upon the field-testing results. In total, 87% of the item sequences in this field test were in the appropriate sequence as written. Recommendations were made to EED with regard to each item sequence. In most cases, it was recommended to simply switch the item order. In some cases, the prompt was adjusted. In others, the student materials were adjusted. The reviewers carefully analyzed each item to determine what made that item more difficult or easier and made recommendations based upon this review. Additional input for the ELOS items will also be gathered via the consequential validity survey, where direct teacher input regarding the assessments will be collected.

The tables below convey the items that were determined to be out of order in terms of item difficulty based upon analysis of field test data. Each content area is demonstrated by grade band, by task number, with the average item scores conveyed, as well as the specific action taken to arrange the items in order of difficulty. For example, in reading, Items 2 and 3 were not in the appropriate order of difficulty (see Non-Ordinal Item Difficulty Stats column). Item 2 had an average level of support of 2.700, while Item 3's average score was 2.767, demonstrating that Item 3 is easier than Item 2. To correct this, both the prompt and student expectation were adjusted to make Item 2 easier.

Reading

| Grade Band | Task | Non-Ordinal Item Difficulty Stats | Response |
|------------|------|-----------------------------------|---|
| 3/4 | 3 | 2-3 (2.700 to 2.767) | ACTION: Change prompt and parenthetical for Task 3, Item 2 |

Writing

| Grade Band | Task | Non-Ordinal Item Difficulty Stats | Response |
|------------|------|-----------------------------------|--|
| 3/4 | 2 | 4-5 (1.525 to 1.609) | ACTION: Change parenthetical for Task 2, Item 5 |
| 3/4 | 3 | 3-4-5 (2.122 to 2.098 to 2.439) | ACTION: Switch the prompts and parentheticals for Task 3, Items 3, 4, & 5. They will now be in the following order: 5, 3, 4 |
| 7/8 | 3 | 3-4 (1.791 to 1.860) | ACTION: Switch the prompts and parentheticals for Task 3, Items 3 & 4 |
| 9/10 | 1 | 1-2 (3.048 to 3.285) | Anomaly; leave intact |
| 9/10 | 3 | 4-5 (1.714 to 1.800) | Difference is minimal; more important to keep task flow intact |

Math

| Grade Band | Task | Non-Ordinal Item Difficulty Stats | Response |
|------------|------|-----------------------------------|--|
| 3/4 | 1 | 3-4 (1.878 to 2.030) | ACTION: Switch the prompts and parentheticals for Task 1, Items 3 & 4 |
| 3/4 | 3 | 4-5 (1.909 to 2.030) | ACTION: Switch the prompts and parentheticals for Task 3, Items 4 & 5 |
| 5/6 | 2 | 3-4 (2.120 to 2.250) | ACTION: Switch the prompts and parentheticals for Task 2, Items 3 & 4 |
| 7/8 | 1 | 3-4 (2.230 to 2.760) | ACTION: Switch the prompts and parentheticals for Task 1, Items 3 & 4 |

| Grade Band | Task | Non-Ordinal Item Difficulty Stats | Response |
|------------|------|-----------------------------------|---|
| 7/8 | 2 | 4-5 (2.436 to 2.410) | ACTION: Switch the prompts and parentheses for Task 2, Items 4 & 5 |
| 7/8 | 3 | 4-5 (2.026 to 2.026) | These are equivalently difficult; leave intact |
| 9/10 | 1 | 3/4-5 (2.172 to 2.276) | ACTION: Switch the prompts and parentheses for Task 1, Items 3, 4, & 5. They will now be in the following order: 5, 3, 4 |
| 9/10 | 2 | 4-5 (1.821 to 2.207) | ACTION: Change the order of the answer choices in the SM for Task 2, Item 5 |
| 9/10 | 3 | 4-5 (2.276 to 2.379) | ACTION: Switch the prompts and parentheses for Task 3, Items 4 & 5 |

Science

| Grade | Task | Non-Ordinal Item Difficulty Stats | Response |
|-------|------|-----------------------------------|---|
| 4 | 1 | 4-5 (1.538 to 1.846) | ACTION: Switch the prompts and parentheses for Task 1, Items 4 & 5 |
| 4 | 2 | 2-3 (2.462 to 2.615) | ACTION: Change the order of the answer choices for Task 2, Item 3 |
| 4 | 2 | 4-5 (1.769 to 1.846) | ACTION: Switch the prompts and parentheses for Task 2, Items 4 & 5 |
| 4 | 3 | 3-4-5 (1.846 to 1.846 to 1.846) | These are all equivalently difficult; no action required |
| 8 | 1 | 3-4 (2.333 to 3.389) | ACTION: Switch the prompts and parentheses for Task 1, Items 3 & 4 |
| 8 | 2 | 3-4-5 (1.889 to 1.667 to 1.833) | ACTION: Switch the prompts and parentheses for Task 2, Items 3, 4, & 5. They will now be in the following order: 3, 5, 4 |
| 8 | 3 | 4-5 (1.824 to 2.33) | ACTION: Switch the prompts and parentheses for Task 3, Items 4 & 5 |
| 10 | 1 | 1-2 (3.667 to 3.667) | These are equivalently difficult; no action required |
| 10 | 1 | 3-4 (2.000 to 2.167) | ACTION: Switch the prompts and parentheses for Task 1, Items 3 & 4 |
| 10 | 2 | 1-2 (4.000 to 4.000) | These are equivalently difficult; leave intact |
| 10 | 2 | 4-5 (1.833 to 1.833) | These are equivalently difficult; leave intact |
| 10 | 3 | 4-5 (2.000 to 2.000) | These are equivalently difficult; leave intact |

CHAPTER 9: ADEQUATE YEARLY PROGRESS

Quality assurance is applied to all Adequate Yearly Progress (AYP) calculations. The original data file is first reviewed by EED for demographic accuracy, most specifically, the Alaska Student Identification numbers (AKSID) and the grade level assignments for each student. This review of submissions for accuracy ensures that only appropriate records used for calculations and that the calculations are performed at the correct grade level. All subsequent AYP calculations are performed by two separate procedures using two separate statisticians. While each statistician performs internal quality checks to ensure the accuracy of their work independently, they also compare files to ensure a 100% match between their results for all records. Historically, agreement has been established for all results beyond the thousandths level. Once a 100% match between the two statisticians is verified, additional, randomized quality assurance checks are performed on the final AYP data file as well as the *Individual Student Reports* (ISRs) generated from the final AYP data file by two additional quality assurance evaluators.

Standard

The first quality assurance evaluator reviewed 17 total student records. The sample represented a variety of districts and regions, but more importantly looked at all possible permutations the syntax used for calculations (this would require only 15 reviews, however, one student record pulled was an ELOS record and one additional file was reviewed for redundancy). Student records were pulled from the AYP file and compared to the original data file. The following verifications were performed by hand without the aid of software.

- 1) Verified student demographics, including student first name, student middle name, student last name, student Alaska student identification number, student grade, student date of birth, student district, student school, and relevant Assessor name.
- 2) Verified all cut scores across reading, writing, math, and science (and the combined ELA cut scores).
- 3) Verified raw scores across reading, writing, math, and science (one content area, per grade band, per student).
- 4) Verified all scaled scores in reading, writing, math, and science (and the combined ELA scaled scores), using the established linear equating formula. This verification included an analysis of current strands and weights, as well as current and former standard deviations and means.
- 5) Verified all AYP performance level assignments.

Individual Student Reports

The first quality assurance evaluator also reviewed 15 additional ISRs, selected to represent all possible permutations of the ISR forms, to ensure that there was 100% match between the final AYP file and what is reported on the ISRs using the same five domains reviewed above. ISRs were reviewed for overall formatting and accuracy, including all data, cut scores, and performance level assignments. In addition, all Performance Level Descriptors (PLDs) were verified. A review was conducted to ensure that all students who

participated in the Alaska Alternate Assessment received ISRs. This review demonstrated that all students were included.

Confirmatory Quality Assurance Review

A second quality assurance evaluator performed the following verifications in order to address potential accuracy concerns established in the previous year. Some of these reviews are purposefully redundant. The second quality assurance evaluator:

- 1) Verified that the reasons not administered are coded appropriately in the final AYP data file.
- 2) Verified that subject area calculations sum correctly and are appropriately matched with the cut score.
- 3) Verified all nulls and zeroes.
- 4) Verified that ISRs contain appropriate data transfer, spelling, headers and footers, layout by grade, and performance level assignment.

Adequate Yearly Progress Report Overview

Adequate Yearly Progress (AYP) results are displayed with each attained score value presented in two different ways for depicting proficiency:

1. Four categories with 1 = Far Below, 2 = Below, 3 = Proficient, and 4 = Advanced
2. Two categories with 0 = Below (with Far Below and Below collapsed) and 1 = Above (with Proficient and Advanced collapsed)

For each table, the data present: (a) the frequency of the score value (Frequency), reflecting the number of students at that score value, (b) the percentage of students (Percent), reflecting the number of students in the grade band with a score value divided by all students taking the alternate assessment, including those with missing score values or in a different grade, (c) the percentage of students (Valid Percent), reflecting the number of students who actually had values divided by only those students with a score value in that grade band, and (d) the percentage of students with score values (Cumulative Percent), reflecting a running accumulation of percentages at/below that specific score value using only students in the grade band. The 'Frequency' and 'Valid Percent' need to be the focus of interpretations.

Reading Adequate Yearly Progress

Reading Grades 3 and 4

In grades 3 and 4, 42.8% of students were proficient (a score of 3), and 17% of students were above proficient (a score of 4). A total of 59.8% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 40.2% of student's received a B (below proficient).

Reading Grades 5 and 6

In grades 5 and 6, 50.3% of students received a score of 3 (proficient), and 1.1% of students received a score of 4 (above proficient) equaling a total of 51.4% of all students achieving proficiency. Receiving scores of 2 (below proficient) or 1 (far below proficient), 48.6% of student's received a B (below proficient).

Reading Grades 7 and 8

In grades 7 and 8, 47.5% of students were proficient (A) receiving either a score of 3 (proficient) or 4 (above proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 52.5% of student's received a B (below proficient).

Reading Grades 9 and 10

In grades 9 and 10, 31.7% of students received a score of 3 (proficient), and 8.4% of students received a score of 4 (above proficient) equaling a total of 40.1% of all students achieving proficiency. Receiving scores of 2 (below proficient) or 1 (far below proficient), 59.9% of student's received a B (below proficient).

*Appendix 7.1 Reading Statistics***Writing Adequate Yearly Progress***Writing Grades 3 and 4*

In grades 3 and 4, 21.9% of students were proficient (3), and 40.4% were above proficient (4). A total of 62.3% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 37.7% of students received a B (below proficient).

Writing Grades 5 and 6

In grades 5 and 6, 36.7% of students received a 3 (proficient), and 22% received a 4 (above proficient). On average, 58.7% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 41.3% of students were given a B (below proficient).

Writing Grades 7 and 8

In grades 7 and 8, 33.8% of students received a 3 (proficient), and 12.5% received a 4 (above proficient). On average, 46.3% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 53.7% of students were given a B (below proficient).

Writing Grades 9 and 10

In grades 9 and 10, 49.7% of students received a 3 (proficient), and 2.4% received a 4 (above proficient). On average, 52.1% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 47.9% of students were given a B (below proficient).

*Appendix 7.2 Writing Statistics***Math Adequate Yearly Progress***Math Grades 3 and 4*

In grades 3 and 4, 27.8% of students were proficient (a score of 3), and 42.4% of students were above proficient (a score of 4). A total of 70.2% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 29.8% of student's received a B (below proficient).

Math Grades 5 and 6

In grades 5 and 6, 33% of students received a score of 3 (proficient), and 42.6% of students received a score of 4 (above proficient) equaling a total of 75.6% of all students achieving proficiency. Receiving scores of 2 (below proficient) or 1 (far below proficient), 24.4% of student's received a B (below proficient).

Math Grades 7 and 8

In grades 7 and 8, 28.7% of students received a score of 3 (proficient), and 22.3% of students received a score of 4 (above proficient) for a total of 51% of students receiving an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 49% of student's received a B (below proficient).

Math Grades 9 and 10

In grades 9 and 10, 28.6% of students received a score of 3 (proficient), and 19% of students received a score of 4 (above proficient) equaling a total 47.6% of students receiving an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 52.4% of student's received a B (below proficient).

*Appendix 7.3 Mathematics Statistics***Science Adequate Yearly Progress***Science Grade 4*

In grade 4, 43.6% of students were proficient (a score of 3), and 30.8% of students were above proficient (a score of 4). A total of 74.4% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 25.6% of student's received a B (below proficient).

Science Grade 8

In grade 8, 24.4% of students were proficient (a score of 3), and 41.9% of students were above proficient (a score of 4). A total of 66.3% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 33.7% of student's received a B (below proficient).

Science Grade 10

In grade 10, 62.8% of students were proficient (a score of 3), and 21.8% of students were above proficient (a score of 4). A total of 84.6% of students were given an A (proficient). Receiving scores of 2 (below proficient) or 1 (far below proficient), 15.4% of student's received a B (below proficient).

Appendix 7.4 Science Statistics

CHAPTER 10: PROGRAM IMPROVEMENT

Program Evaluation

The Alaska Alternate Assessment undergoes ongoing and multiple-level evaluation of effectiveness and reliability. In addition to a *Survey of Consequential Validity*, DRA and EED analyze the use of the technical components of the training and score reporting system, verify the effectiveness of training on scoring consistency, analyze the use and appropriateness of accommodations employed in administering the assessments, and review help desk calls for areas requiring additional training.

Summary of Consequential Survey

Assessors are asked annually to complete a survey regarding the Alternate Assessment, their instruction and curriculum, and information about themselves.

Training and Qualifications

Approximately 40% of the Assessors who responded to the survey had between six and fifteen years' experience as educators. About 23% of the respondents had more than sixteen years' experience.

Instructional Relevance

One hundred eight Assessors believe that they do not teach differently as a result of the Alternate Assessment; 99 believe that they do.

Revised ELOS Documents & Assessment

This year's Consequential Survey included questions regarding Assessors' opinions of the revised ELOS Scoring Protocol, Student Materials, and "fit" to the students engaged in the ELOS administration.

A majority of Assessors answered the question "The Alternate Assessment is accessible to my students." as Agree or Strongly Agree (201 and 34). Additionally, 203 Assessors believe that their students who participate in the Alaska Alternate Assessment are improving in their academic skills, while only 23 didn't believe that.

Revised ELOS Administration Format and Procedures

Of particular interest to EED and the test vendor were responses to survey questions regarding the revised ELOS test documents and protocols. Answers of "NA" were excluded from the totals and percentage calculations.

A little more than half of the Assessors responding (54%) to the statement: "The students I teach who are taking the Alternate Assessment are likely to meet the academic standards assessed by the Alternate Assessment" responded as Agree or Strongly Agree.

Two hundred three of two hundred forty seven (82%) of the Assessors believe that “The students I teach who are taking the Alternate Assessment are improving in their academic skills” responded “Agree” or “Strongly Agree.”

Eighty-two percent of the Assessors who responded believed that the Early Entry Points expectations are helpful in planning instruction, while only forty-nine percent of respondents believe that the “new ELOS test items are a good match to my students’ academic skills.

Fifty-nine percent of respondents agreed or strongly agreed that the information received from the ELOS administration is meaningful.

Sixty-four percent of the respondents felt that the new format of the ELOS test documents was beneficial to them.

Eighty-two percent of Assessors who answered believed that it is important to use classroom materials when assessing students who require the ELOS administration.

Teacher Demographics and Experiences

Almost half of Assessors have earned a Master’s degree or higher (47%), and 60% have earned a Special Education certificate, with 40% holding a General Education certificate.

Appendix 10.1a Consequential Validity Survey

Appendix 10.1b Consequential Validity Survey Summary

Appendix 10.1c Consequential Validity Survey Written Responses

Recommendations for Future Consideration

Technological Improvements

Representatives from Dillard Research Associates met with key personnel in Alaska’s Assessment division to plan for the 2012-2013 testing year. The team agreed that no technological changes needed to be made. Two items that involved the web-based training or reporting sites will be implemented, but essentially document changes and uploads.

Recommendations for Training

During the 2011-2012 school year, Assessors participated in one or more training venues:

Face-to-face training led by DRA and EED:

- New Mentor Training
- All Mentor Training

Face-to-face training led by QTs: In-district trainings led by QTs

In addition, all Assessors participated in self-paced, individual training through the ak.k12test.com website, participated in online proficiency testing, entered data into the Data Entry site (including information on accommodations used), and accessed the Help Desk for a variety of issues.

Each of these venues produced data that are useful to record and analyze to determine possible impacts on training in the 2011-2012 school year.

Training Recommendations from EED-led trainings

Twenty AITs participated in the September 2011 New Mentor Training. The *New Mentor Rating Proficiency and Protégé Scoring Accuracy Report* (Appendix 3.2) provides the following recommendations related to noted patterns of errors:

"Future trainings should over-emphasize the need for reviewers to ensure that all protégés have recorded the student's responses, primarily, and verified that the correct score points have been awarded, secondarily" (page 6).

Training Recommendations from Accommodations Used

The accommodation "Hand over hand assistance" continues to be recorded by Alaska Alternate Assessment QAs and QTs. Under some circumstances, this accommodation may serve as a modification. A modification is an adaptation that alters the construct being tested and invalidates the question where the modification was employed.

Recommendations:

Reinforce in New Mentor and Annual Mentor training the differences between modifications and accommodations. Enhance the ak.k12test.com training site to more fully examine this issue. Follow up with the Districts in question.

Training Recommendations from the Writing Study

The writing scoring rules require no modifications and the system is now stable; however, the field could benefit from additional guidance and training in specific areas of writing scoring raw scores in CWS, L, and I&O. Few errors were noted in the calculation of the maximum number of points possible in CWS, with the exception of scoring multiple sentences and raters appearing to "forget" the extra sequence between sentences. Few errors were noted in terms of the Assessors' abilities to calculate the percentages and points earned, once the raw score points were determined. Legibility scoring training needs to emphasize the perspective that a person who is unfamiliar with the writer needs

to be able to decode the text. More detailed review of the I & O rubric is also required for training purposes, as R1s were not applying the rubric in a consistent manner.

In addition to the data review, evaluators also collected anecdotal evidence as writing scoring protocols and student materials were reviewed. Information gathered during this anecdotal review process will be used to develop an addendum for the Writing Scoring Manual. The addendum will provide new examples, scoring, and explanations of the scoring procedures applied. These examples will include sentences as well as stories and will be selected such that they support the development of Assessor skills in the areas of rounding, accuracy, and self-consistency, and Assessor knowledge in the areas of scoring: 1) CWS with provided words, 2) multiple sentences, 3) internal capitalization, and 4) I & O. In addition, more attention related to use and implementation of the 3X3 rule for determining ELOS administration will be included in next year's trainings.

Training Recommendations from Proficiency Testing

The greatest number of Assessors who required three or more attempts to pass a proficiency test occurred in the subject areas of math and writing.

Recommendation: Analyze the test items that were most frequently failed for content. Enhance face-to-face and online training around those content issues. Appendix 4.1 will serve as the basis of that analysis.

Training Recommendations from HelpDesk Questions

During the 2012 testing window, DRA's HelpDesk operator (Sevrina Tindal) answered seventy-seven inquiries from Alaska Assessors of the Alternate Assessment. Their queries represented thirteen topics and are summarized in Appendix 10.2.

Appendix 10.2 HelpDesk Log

Most of the HelpDesk calls related to operational or procedural questions, and were answered quickly. Follow up assistance was provided as necessary. The complete report ("1112_AKHelpdeskLog-VF.xls") is located on the Secure Server.

The four categories with the greatest number of calls (marked with an * in the table) are analyzed below.

| Issue | N |
|-------------------------------------|----|
| 2011 Individual Student Report | 1 |
| AK website logon difficultes | 1 |
| Data Entry * | 9 |
| District Account Management | 5 |
| District Training alert | 2 |
| Online Training * | 29 |
| Proficiency Testing | 3 |
| Secure Tests | 4 |
| Test Administration * | 8 |
| Testing Window | 1 |
| Unofficial Student Report | 1 |
| Writing Study | 1 |
| Learner Characteristics Inventory * | 11 |

- **Online Training**

Of the 29 questions received regarding Online Training, seven related to Refresher status due to returning QAs requesting new logins, five related to assisting with new passwords; the remainder reflected various issues around using the online training system, enrolling new users, and individual computer issues at the district.

- **Learner Characteristic Inventory**

Of the 11 questions received regarding the Learner Characteristic Inventory, seven related to an issue discovered on the first day, wherein Survey Monkey appeared to track the IP address of computers submitting responses, and allowed only one response per computer. This prevented Assessors from responding to the LCI with multiple students. This issue was quickly repaired by EED. The remaining questions related to receiving confirmations that the LCI was completed for students, and one question about a "frozen" survey that didn't permit the user from moving beyond 7% completion.

- **Data Entry**

Of the nine help desk calls related to Data Entry, four related to issues about deleting a student from an Assessor's case load; three related to confusion around the February 15, 2012 goal to enter student data into the system; one related to an Assessor who forgot to enter a student's ELOS scores and one related to a question around District ID numbers.

- **Test Administration**

Eight clarifying questions were received regarding a variety of issues around test administration rules, including questions about ELOS administration for a student who is more skilled than the ELOS assessment but not proficient in the Standard administration, questions around the NA-I category counting toward the 3 X 3 rule and the use of NA-I for students who are non-verbal, a question around the proper way to administer a Large Print version, and questions about test materials disposal at the end of the testing window.

- **Test Document errata**

There were no instances of discrepancies or errata on the 2012 test documents.

Recommendations:

With the exception of the Learner Characteristic Inventory, the other categories with frequent queries should be addressed in training, both by enhancing the web site training sections for those particular issues, and by highlighting the issues and solutions in face-to-face training.