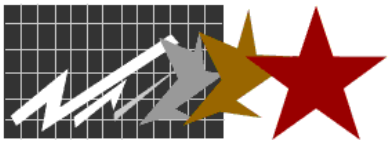

Dillard Research Associates

Alaska Alternate Assessment

2014 - 2015

Technical Report

Dillard Research Associates
June 30, 2015

Alaska Alternate Assessment Technical Report

Table of Contents

EXECUTIVE SUMMARY 1

CHAPTER 1: BACKGROUND OF THE ALASKA ALTERNATE ASSESSMENT 3

Overview 3

History of Previous Program *3*

Current Program Overview 3

Reasons for Current Approach *3*

Roles of Contractor, Department, and Others *4*

Summary of Current Program 4

Description of Program *4*

Description of Students Served *4*

Description of How Scores Are Used *5*

Significant Changes Since Previous Technical Report 5

Organization of Technical Report 5

CHAPTER 2: TEST DESIGN AND ITEM/TASK DEVELOPMENT 7

Overview 7

Description of ExGLEs and their relationship to GLEs *8*

Test Specifications and Blueprint 9

Description of Test Specifications (DOTS) *9*

Process of Establishing Test Specifications *9*

Item Content Test Blueprint and Item Specifications *10*

Proficiency Level Descriptor Development *10*

Cut Scores *10*

Item/Task Development 11

Item Writing, including Scoring Guides *11*

Expanded Levels of Support (ELOS) *11*

Reduction in Complexity, Depth, and Breadth *11*

Bias and Sensitivity Review *13*

Test Design and Development 14

Representation and Functionality *14*

Psychometric Guidelines for Selecting Items/Tasks for Item/Task Bank *14*

Current Test Construction 15

Construction of the Operational Forms *15*

Test Development Timeline *15*

Internal Review of the Items and Forms *16*

CHAPTER 3: TEST ADMINISTRATION PROCEDURES 17

Overview 17

Student Population Tested 17

Standard Administration With or Without Accommodations *17*

Standard Administration With or Without Accommodations AND Then Switched to the ELOS *18*

ELOS Administration *19*

Accommodations 19

Test Administrators 20

Mentor Responsibilities....20
Materials....21
Test Administrator Training....21
Annual Mentor Training....*21*
Webinars....*22*
Online Training....*22*
Refresher Training and Testing....*23*
Security....*23*

CHAPTER 4: SCORING....24
Overview....24
Quality Control of Scoring....24
Procedures....*24*
Data Entry....24
Analysis of Accommodations Used....*27*

CHAPTER 5: STANDARDS VALIDATION....28

CHAPTER 6: REPORTING....29
Overview....29
Reporting Student Results....30
DRA Secure Reporting Website....30

CHAPTER 7: TEST VALIDITY....31
Overview....31
Validity....31
Reliability....32
Total Test Reliabilities (All Students)....33
Science Reliability....*33*
Item Analysis of ELOS Administration....*33*

CHAPTER 8: DESCRIPTIVE STATISTICS....36
Strand, Task, and Item Difficulties....36
Strand Difficulties in Science (Standard, No ELOS)....*36*
Science Strand Difficulties....*36*
Science Task Difficulties....*37*
Science Item Difficulties....*37*

CHAPTER 9: ANNUAL MEASURABLE OBJECTIVES....41
Standard....*41*
Individual Student Reports....*41*
Confirmatory Quality Assurance Review....*41*
Annual Measurable Objectives Report Overview....*42*
Science Annual Measurable Objectives....*42*

CHAPTER 10: PROGRAM IMPROVEMENT....43
Program Evaluation....43
Summary of Training Evaluations....43
Summary of Consequential Survey....43
Training and Qualifications....*44*
Accessibility and Impact....*44*
Summary of Help Desk Queries....44
Recommendations for Future Consideration....45

Technological Improvements *45*
Recommendations for Training 45
Training Recommendations from Consequential Validity Survey Responses *45*
Training Recommendations from HelpDesk Questions *45*

APPENDICES

Appendix 1.1 Web Changes Handout 5
Appendix 2.1 Science Proficiency Level Descriptors 10
Appendix 2.2 Linguistic Complexity & Bias and Sensitivity Review 13
Appendix 2.3 Quality Assurance Manual 16
Appendix 3.1a Annual Mentor Training Attendees 21
Appendix 3.1b Annual Mentor Training Agenda 21
Appendix 3.1c IEPs linked to DLM Essential Elements 21
Appendix 3.2 January Annual Mentor Training Agenda 22
Appendix 3.3a Getting Started Webinar 22
Appendix 3.3b Pre-Test Webinar 22
Appendix 3.3c Post-Test Webinar 22
Appendix 3.4 AKAA 2015 Training Site Table of Contents 22
Appendix 3.5 Test Security Agreement 23
Appendix 3.6 Test Site Security 23
Appendix 4.1 Accommodations Used Summary 27
Appendix 6.1 Unofficial and Official Individual Student Report Matrix 30
Appendix 6.2a Educator and Parent Guides to Reading, Writing, and Math Reports 30
Appendix 6.2b Educator and Parent Guides to Science Report 30
Appendix 6.3 Reporting Website Manual 30
Appendix 7.1 Science Statistics 42
Appendix 10.1a October Training Evaluation Summary 43
Appendix 10.1b January Training Evaluation Form 43
Appendix 10.2 Consequential Validity Report 44
Appendix 10.3 2015 HelpDesk Log 44

EXECUTIVE SUMMARY

The purpose of this *Technical Report* is to record the administration and reporting of the 2014-2015 Alaska Alternate Assessment. For this year, Dillard Research Associates (DRA) was responsible for the Science Alternate Assessment only, administered to eligible students in grades 4, 8, and 10.

The domains 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 areas of training, administration, scoring, and reporting related to the Alaska Alternate Assessment (AKAA). In addressing technical documentation, we first present content evidence, then reliability, then descriptive statistics and Annual Measurable Objective (AMO) calculations that are used to inform the Alaska State Performance Index (ASPI).

In the end, both procedural and empirical evidence support 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 AKAA.

Chapters 7 and 9 share the same appendix. Chapter 8 includes strand, task, and item difficulty statistics within the body of the technical report.

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

AMO

- Participation descriptive statistics at the total test level for each grade
- Score descriptives for total tests, including the frequencies of each score

Test Strand Descriptive Statistics

- Test Strand descriptives for each grade (4, 8, and 10)
- Strand descriptives for each grade (4, 8, and 10)

Task Item Descriptive Statistics

- Operational task descriptives for each grade (4, 8, and 10)
- Task descriptives for each grade (4, 8, and 10)
- Task item descriptives for each grade (4, 8, and 10)

Reliability

- Item reliabilities for each grade (4, 8, and 10)

The AMO tables depict percentages of students participating, the number of students at each score value, and the total sum.

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 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 statistics also are presented.

CHAPTER 1: BACKGROUND OF THE ALASKA ALTERNATE ASSESSMENT

Overview

The 2014–2015 Alaska Alternate Assessment in Science represented an equivalent form test to the 2012–2013 Alternate Assessment. This version of the assessment is referred to as "Form B."

History of Previous Program

In 2005, a Reliability and Validity study was conducted by Dr. Gerald Tindal. The study conclusion was that the State of Alaska's Student Portfolio system needed revision in order to meet technical quality requirements set by the No Child Left Behind legislation. As a result of the department's Request for Proposals process, 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). Teams of content experts created the Proficiency Level Descriptors (PLDs).

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 (NCEO), 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 AKAAs are standardized performance tasks administered and scored by Assessors who undergo a multi-step qualification process. The U.S. ED Title 1 Final Assessment System Peer Review process has approved Alaska's current system of assessing students with significant cognitive disabilities.

Roles of Contractor, Department, and Others

The contractor, Dillard Research Associates (DRA), served the Alaska Department of Education and Early Development (EED) since 2005 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. Prior to this school year, DRA produced the Alternate Assessments in all subject areas (Reading, Writing, Mathematics, and Science). For the 2014-2015 testing year, Dynamic Learning Maps produced the Assessments for Reading, Writing, and Mathematics, while DRA produced the Science Assessment only.

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

Summary of Current Program

Description of Program

The Science Alternate Assessment is composed of standardized performance tasks administered and scored by Assessors who undergo a multi-step qualification process.

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 (QAs) or Qualified Trainers (QTs) during the test window
- A data entry and reporting portal
- A secure reporting site for district access to individual student reports

Description of Students Served

The Science Alternate Assessment is administered to students with significant cognitive disabilities in grades 4, 8, and 10 and measures 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.

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. These scores form the basis for the AMO report for these students. Official student reports that have demographic information checked for accuracy and include assigned proficiency levels were made available to districts on May 18, 2015 via the District Test Coordinators at the secure DRA Web Reporting System.

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

Significant Changes Since Previous Technical Report

Improvements were made to several key areas of the AKAA for the 2014-2015 testing window, primarily related to the dual testing system represented by DLM (Reading, Writing, and Mathematics) and DRA (Science).

Training and Continuous Improvement

- Developed online training for QAs seeking to become QTs, including protocols for district QTs to assist, monitor, and evaluate. In districts without a QT, DRA served in that role, and
- Included training related to accessing and completing steps required to administer DLM assessments.

Appendix 1.1 Web Changes Handout

Organization of Technical Report

The 2014 *Technical Report* is organized around ten broad topics, with detailed appendices referenced where appropriate. The *Technical Report* serves as a narrative description of the activities and results of the 2014-2015 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:

1. Background of the Alaska Alternate Assessment
2. Test Design and Item/Task Development
3. Test Administration Procedures
4. Scoring
5. Standards Validation
6. Reporting

7. Test Validity
8. Descriptive Statistics
9. Annual Measurable Objectives (AMOs)
10. Recommended Program Improvements

CHAPTER 2: TEST DESIGN AND ITEM/TASK DEVELOPMENT

Overview

The NCEO 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 not included in district or state assessment systems. 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 AKAA currently provides test scores that are used to determine Annual Measurable Objectives (AMOs), which then feed Alaska School Performance Index (ASPI) accountability ratings.

The original design of the AKAA, 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 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 focused on measuring reading, writing, mathematics, and science. The current AKAA 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 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 Performance Level Descriptors (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.

In June 2012, the Alaska State Board of Education adopted new standards for reading, writing, and mathematics for grades K-12. These standards are equal in rigor to the common core state standards (CCSS). The Alaska Education and Early Development division (EED) also joined the Dynamic Learning Maps (DLM) consortium of states working together to develop and administer a new AA-AAS for students with significant cognitive disabilities. The DLM assessment will address the new Alaska State Standards (AKSS), while the current AKAA addresses the ExGLEs.

Test Specifications and Blueprint

Description of Test Specifications (DOTS)

Descriptions of Test Specifications for the 2014–2015 Alternate Assessment are Excel spreadsheets that define all aspects of each item used in all test materials. In addition to items used in the 2014–2015 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 2015 testing window matches the Form B test administered in 2012. Science is not weighted, so is not included in this appendix.

Proficiency Level Descriptor Development

Prior to the adoption of the new AKAA, 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 2.1 Science 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:

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

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 2014 test window. The 2014-2015 AKAA test documents matched the test documents deployed in 2012-2013, and are referred to as "Form B."

Expanded Levels of Support (ELOS)

Between June and September 2012, the ELOS test documents were substantially improved. Previously, one set of documents covered the ELOS administration for all four grade bands in reading, writing, mathematics and science, respectively. In the spring of 2012, new ELOS assessments were developed at each grade band separately, each composed of three tasks with five items that must all be administered. 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. The test documents used in 2015 are identical to those used in the 2013 and 2014 administrations.

Assessors rate the level of support needed to bring the student to success on a 4-pt scale (1 = full physical support, 2 = partial physical support, 3 = visual, verbal, and/or gestural prompts, 4 = student completes task independently).

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. Karvonen and Almond conducted an alignment study in 2007; 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.

The ELOS items developed last year employed 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.

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 2.2 Linguistic Complexity & 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.

Current Test Construction

The 2008-2009 AKAA 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 2012-2013, 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 2014-2015 (Form B) are operational items.

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

1) AKAA Test (2007-2008, 2008-2009)
2) 1/2 (a) AKAA test + 1/2 (a) FT (FORM A, 2009-2010)
3) 1/2 (b) AKAA test + 1/2 (b) FT (FORM B, 2010-2011)
4) 1/2 (a) AKAA test + 1/2 (b) FT
5) 1/2 (b) AKAA 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 AKAA (Form A and Form B, numbers 2 and 3 above).

1) Form A, 2011-2012 and 2013-2014 (described in #2 above)
2) Form B, 2012-2013 and 2014-2015 (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 AKAA. 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 2013 test year.

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

Internal Review of the Items and Forms

DRA maintains iterative internal and external quality assurance procedures and reviews 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 2.3 Quality Assurance Manual

CHAPTER 3: TEST ADMINISTRATION PROCEDURES

Overview

The AKAA 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, December 2014 edition, located at: https://ak.k12test.com/files/2014_December_Participation_Guidelines.pdf?1422989734

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

The AKAA 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.

Including Student Participation and Performance

Students taking the AKAAs (including ELOS) can be counted in their school and district for Annual Measurable Objectives (AMOs) in the areas of performance and participation. Individual student scores are calculated and assigned a proficiency level: Advanced, Proficient, Below Proficient, or Far Below Proficient. The ELOS items receive scores, but the proficiency level is Far Below Proficient. All students receive individual student reports.

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 AKAA. 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 AKAA allows for accommodations to be utilized during test administration. The student's IEP team determines accommodations for the student. 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.

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 level of support needed to bring them to success on the item.

Accommodations

The AKAA allows accommodations to be utilized during test administration. The student's IEP team determines accommodations for each student.

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_2013.pdf

Test Administrators

Only school personnel may administer the AKAA. 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 Qualified Mentor-Trainer (QT). Districts with more than one QT appoint one person to serve as the Lead QT in interactions with EED. A Mentor-in-training first must meet all of the training requirements to become certified as a QA.

A mentor-in-training completes new mentor training online and attends the annual mentor training. To be upgraded to Qualified Mentor-Trainer, a participant must:

- 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
    - Upgrading protégé status from AIT to QA after the protégé has produced corrected scoring protocols to the qualifying level

After meeting qualifications, QTs become certified and have their status upgraded by their district's QT. If the district does not have a QT, DRA evaluated the work and upgraded the QA's status to QT.

Ongoing requirements to continue as a QT:

- Hold a QT Certificate
- Attend any required refresher trainings
- Refresh proficiency annually to maintain access to online system
- Sign Test Security Agreements annually, keeping one copy and filing one with the District Test Coordinators (DTCs)

Mentors have access to online reports to track the district test administrators' progress through training, update user status to QA when appropriate, track progress toward entering student demographic information and toward completion of assessment administrations, and track any Assessors who have not completed student assessments and submitted the scores to EED 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 and are embargoed until the Friday prior to the test window opening. The training pages and support materials for training are available to all registered users.

Test Administrator Training

The purpose of the AKAA Mentor Program is to prepare district level trainers who train district personnel in correct test administration procedures for the AKAA. 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 AKAA. 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 text based training sections and 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 questions related to general test administration.

Annual Mentor Training

The additional responsibilities of a QT necessitate additional training, which was held October 6, 7, and 8, 2014 in Anchorage. This training provided more in-depth information on the changes to the 2014-2015 AKAAs and Secure website, including training tips to the QTs; requirements to be proficient Assessors of the new DLM assessments in Reading, Writing, and Mathematics; and changes in the ak.k12test.com training and testing site.

The appendix contains information and handouts related to the DRA training sessions on the first day of training. To access documents and topics addressed by DLM on the second day of training, please visit: http://dynamiclearningmaps.org.

Appendix 3.1a Annual Mentor Training Attendees

Appendix 3.1b Annual Mentor Training Agenda

Appendix 3.1c IEPs linked to DLM Essential Elements

In addition to the traditional Fall Annual Mentor Training, a second training was provided to Mentors to prepare them for the instructionally-embedded testing system developed by DLM. This training was held January 21 and 22, 2015 in Anchorage. Alaska's Education and Early Development department has hosted the training materials on their website at: http://education.alaska.gov/tls/assessment/MentorSPEDTrainingMaterials_Jan2015.html

Appendix 3.2 January Annual Mentor Training Agenda

Webinars

Three webinars were held throughout the year. The first, called the "Getting Started" webinar was offered in September to prepare Mentors with pre-training homework prior to attending the October Annual Mentor Training.

Appendix 3.3a Getting Started Webinar

On March 11, 2015, DRA, DLM, and EED hosted a webinar for QTs of the updates to the AKAA website. The webinar served to remind QTs of the changes to the AKAA system, with a focus on the new DLM assessments. The agenda and attendance report are included in the appendix.

Appendix 3.3b Pre-Test Webinar

On May 6, 2015, DRA, DLM, and EED hosted a webinar to address the steps required for closing out the testing year for the DLM assessments and a report on the final Science Alternate Assessment submissions. In addition, QTs were informed of changes to the Assessment system scheduled for 2015-2016, particularly the assumption of the Science assessment by the Kansas testing group, the resulting exiting of the DRA assessment team, and the June 30, 2015 retirement of Aran Felix. DRA assisted in the delivery of the webinar. The agenda, handouts, and attendance report are included in the appendix.

Appendix 3.3c Post-Test Webinar

Online Training

All Assessors must complete the online training through the ak.k12test.com site. After participating in training through the ak.k12test site on all aspects of administering, scoring, and data entry for the AKAA, Assessors-in-Training (AITs) participate in proficiency testing. Each of two training areas is tested with a 20-question multiple-choice test (Administration 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 AKAA 2015 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.

Refresher Training and Testing

Because this was the first year to include the DLM assessment system, all Assessors were required to complete all training topics. There were no reduced "refresher" requirements for returning Assessors.

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.

Prior to the opening of the Test Window, all QAs and QTs are required to sign test security agreement and submit this document to their District Test Coordinator.

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. 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. Following the close of the test administration window, all testing materials must be shredded and electronic versions removed from computer hard drives.

The Test Security Agreement is available in the appendix.

Appendix 3.5 Test Security Agreement

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

Appendix 3.6 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 will administer Alternate Assessments to their students participate in rigorous, self-paced training through the ak.k12test.com online training system. This system also included a section designed to lead Assessors through the training requirements of the DLM assessment system. At the conclusion of online training, participants complete an online proficiency test. Participants must earn 80% or higher in each subject area (administration 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	1st attempt	2nd attempt	3rd attempt	4th attempt	5 or more attempts
Administration	258	26	5	NA	NA
Science	285	5	0	1	NA

Data Entry

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

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 the Science Assessment or given a reason why the test wasn't administered. Scores not submitted by the close of the testing window are invalidated.

Following are the "reasons not tested" that a QA or QT would choose to alert DRA and EED about why a student will not be tested. This information is located in the Data Entry section of the online assessment system.

1. IEP Change This code is selected for students who have an IEP change indicating they are no longer eligible to take the AKAA 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 AKAA.

2. Late Entry This code is entered for students who enter the district from out of state or from a private school after the AKAA 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 2010-2013 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 AKAA.

For the 2014-2015 assessment, as in 2012-2013, 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

For option D, Assessors were instructed to choose all tests administered in one session: [checkboxes, allow one or all to be chosen]: Reading, Writing, Math, and Science

"Breaks" means that the student was provided frequent breaks during testing.
"Multiple sessions" means that the test was administered over several days.

The table below presents the frequencies of timing and scheduling accommodations:

Choice	Science
A	73
B	175
C	7
D	3

In addition, teachers of students eligible for the AKAA are encouraged to use the AKAA 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 AKAA testing format prior to administration of the official AKAA.

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 him/herself 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 2014-2015 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. It is important to understand that these numbers will not match because students typically received more than one accommodation.

- In **Science**, there were 12 types of accommodations made to the test administration with 7 types of accommodations made for student responses and 6 types of accommodations made to the test materials. Fifty-eight students received one or more accommodations in science.

Appendix 4.1 Accommodations Used Summary

ELOS Scoring

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 AKAA reports.

All tables and analyses are presented for science results in a standardized layout format.

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 AMO calculations are completed. The Official Student Report reports a student's proficiency level relative to the Science Extended Grade Level Expectations (ExGLE).

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

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. Though DRA was not responsible for Reading, Writing, and Math Assessments in the 2014–2015 testing year, the Educator and Parent Guides for those subject areas were updated and made available for IEP teams discussing previous years' results.

Appendix 6.2a Educator and Parent Guides to Reading, Writing, and Math Reports

Appendix 6.2b Educator and Parent Guides to Science Report

DRA Secure Reporting Website

Official Individual Student Reports were made available to each district's District Test Coordinator and Qualified Mentor-Trainers on May 16, 2015. Reports are downloaded from the secure Reporting Website at akreports.k12test.com. 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 AMO calculations, test strand descriptive statistics, task descriptive statistics, task item descriptive statists 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 the Annual Measurable Objective (AMO) 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 AMO) 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 AMO; however their scores are reported as Far Below Proficient for AMO performance reporting.

Total Test Reliabilities (All Students)

Science: Grade 4 (.921 for 85 students taking 24 items), grade 8 (.848 for 67 students taking 24 items), and grade 10 (.845 for 66 students taking 24 items).

Appendices 7.1 Science Reliability Statistics

Note - only tasks with at least 4 items are included

Science Reliability

Science Grade 4

Task Name	Cronbach's Alpha
1.4: Concepts of Physical Science	.770
2.4: Concepts of Life Science	.749
3.4: Concepts of Earth Science	.714
4.4: History and Nature of Science, Science and Technology	.725

Science Grade 8

Task Name	Cronbach's Alpha
1.8: Concepts of Physical Science	.609
2.8: Concepts of Life Science	.651
3.8: Concepts of Earth Science	.587
4.8: Science and Technology	.592

Science Grade 10

Task Name	Cronbach's Alpha
1.10: Concepts of Physical Science	.609
2.10: Concepts of Life Science	.548
3.10: Concepts of Earth Science	.603
4.10: Science and Technology	.678

Item Analysis of ELOS Administration

The Science ELOS were reviewed to ensure that the assessments continued to function 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 continues to function well, with most items arranged in order of difficulty.

The ELOS assessments reflect an appropriate range of item difficulties, with average item difficulties ranging from 1.58 to 3.33 in Science. All item difficulties are reported below. Average item difficulty across all grades is 2.60 in science. These results are similar to those garnered over the past three years on the ELOS assessments.

ELOS Science Grade 4

Task Number	Item Number	Average Score (1-4)
Task 1	1	3.05
Task 1	2	2.79
Task 1	3	1.84
Task 1	4	1.58
Task 1	5	1.63
Task 2	1	2.79
Task 2	2	2.37
Task 2	3	2.21
Task 2	4	2.05
Task 2	5	2.26
Task 3	1	2.95
Task 3	2	2.84
Task 3	3	1.89
Task 3	4	1.68
Task 3	5	2.11
TOTAL		**34.05**

ELOS Science Grade 8

Task Number	Item Number	Average Score (1-4)
Task 1	1	3.00
Task 1	2	2.89
Task 1	3	2.44
Task 1	4	2.56
Task 1	5	1.88
Task 2	1	3.00
Task 2	2	2.78
Task 2	3	2.78
Task 2	4	2.56
Task 2	5	2.89
Task 3	1	3.22
Task 3	2	3.33
Task 3	3	3.11
Task 3	4	3.00
Task 3	5	2.22
TOTAL		**41.44**

ELOS Science Grade 10

Task Number	Item Number	Average Score (1-4)
Task 1	1	3.00
Task 1	2	2.96
Task 1	3	2.61
Task 1	4	2.71
Task 1	5	2.41
Task 2	1	3.26
Task 2	2	3.14
Task 2	3	2.61
Task 2	4	2.64
Task 2	5	2.27
Task 3	1	3.22
Task 3	2	3.00
Task 3	3	2.39
Task 3	4	2.55
Task 3	5	2.50
TOTAL		**40.35**

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 "Science Task Descriptive Statistics (Grade Band 3/4)" refers to the descriptive statistics for science 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 Science (Standard, No ELOS)

The tables provided below elaborate the strand difficulties for science 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). Low values are difficult and high values are easy.

Science Strand Difficulties

The most difficult strands are Concepts of Physical Science in Grade 10 with a 49% success rate, and Concepts of Earth Science in Grade 8 with a 52% success rate. The easiest strand is Science and Technology in Grade 8, with a 73% success rate.

Strand Name	*p*
Concepts of Physical Science 4	.58
Concepts of Life Science 4	.61
Concepts of Earth Science 4	.62
Science and Technology 4	.56
Concepts of Physical Science 8	.61
Concepts of Life Science 8	.66
Concepts of Earth Science 8	.70
Science and Technology 8	.73
Concepts of Physical Science 10	.49
Concepts of Life Science 10	.56
Concepts of Earth Science 10	.52
Science and Technology 10	.66

Science Task Difficulties

Science task difficulties range from .49 to .73.

Science Tasks Grade 4

The most difficult task in Grade 4 is History and Nature of Science, Science and Technology, with a success rate of approximately 56%. The easiest task in Grade 4 is Concepts of Earth Science, with a success rate of approximately 62%.

Task Name	*p*
Concepts of Physical Science	.58
Concepts of Life Science	.61
Concepts of Earth Science	.62
History and Nature of Science, Science and Technology	.56

Science Tasks Grade 8

The most difficult task in Grade 8 is Concepts of Physical Science, with a success rate of approximately 61%. The easiest task in Grade 8 is Science and Technology, with a success rate of approximately 73%.

Task Name	*p*
Concepts of Physical Science	.61
Concepts of Life Science	.66
Concepts of Earth Science	.70
Science and Technology	.73

Science Tasks Grade 10

The most difficult task in Grade 10 is Concepts of Earth Science, with a success rate of approximately 52%. The easiest task in Grade 10 is Science and Technology, with a success rate of approximately 66%.

Task Name	*p*
Concepts of Physical Science	.49
Concepts of Life Science	.56
Concepts of Earth Science	.52
Science and Technology	.66

Science Item Difficulties

Science item difficulties range .81 to 1.93. All items are out of a maximum of 2.0 points.

Science Items Grade 4

The most difficult item in Grade 4 is 4.4, Item 5, with a success rate of approximately 36%. The easiest items in Grade 4 are 2.4, Item 3, 3.4, Item 2, and 3.4, Item 6, with success rates of approximately 78%.

Item	*p*
Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_1	.72
Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_2	.61
Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_3	.65
Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_4	.47
Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_5	.72
Science_Grade_4_Task_1.4_Concepts_of_Physical_Science_Item_6	.47
Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_1	.70
Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_2	.64
Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_3	.78
Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_4	.65
Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_5	.49
Science_Grade_4_Task_2.4_Concepts_of_Life_Science_Item_6	.57
Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_1	.63
Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_2	.78
Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_3	.69
Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_4	.56
Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_5	.40
Science_Grade_4_Task_3.4_Concepts_of_Earth_Science_Item_6	.78
Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_and_Technology_Item_1	.64
Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_and_Technology_Item_2	.59
Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_and_Technology_Item_3	.52
Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_and_Technology_Item_4	.77
Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_and_Technology_Item_5	.36
Science_Grade_4_Task_4.4_History_and_Nature_of_Science,_Science_and_Technology_Item_6	.60

Science Items Grade 8

The most difficult item in Grade 8 is 1.8, Item 5, with a success rate of approximately 45%. The easiest item in Grade 8 is 3.8, Item 6, with a success rate of approximately 93%.

Item	*p*
Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_1	.51
Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_2	.82
Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_3	.48
Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_4	.71
Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_5	.45
Science_Grade_8_Task_1.8_Concepts_of_Physical_Science_Item_6	.80
Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_1	.77
Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_2	.65
Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_3	.59
Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_4	.56
Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_5	.82
Science_Grade_8_Task_2.8_Concepts_of_Life_Science_Item_6	.77
Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_1	.68
Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_2	.79
Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_3	.73
Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_4	.65
Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_5	.59
Science_Grade_8_Task_3.8_Concepts_of_Earth_Science_Item_6	.93
Science_Grade_8_Task_4.8_Science_and_Technology_Item_1	.92
Science_Grade_8_Task_4.8_Science_and_Technology_Item_2	.50
Science_Grade_8_Task_4.8_Science_and_Technology_Item_3	.72
Science_Grade_8_Task_4.8_Science_and_Technology_Item_4	.81
Science_Grade_8_Task_4.8_Science_and_Technology_Item_5	.79
Science_Grade_8_Task_4.8_Science_and_Technology_Item_6	.83

Science Items Grade 10

The most difficult item in Grade 10 is 3.10, Item 6, with a success rate of approximately 24%. The easiest items in Grade 10 are 4.10, Items 5 and 6, both with success rates of approximately 82%.

Item	*p*
Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_1	.56
Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_2	.45
Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_3	.62
Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_4	.50
Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_5	.58
Science_Grade_10_Task_1.10_Concepts_of_Physical_Science_Item_6	.45
Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_1	.69
Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_2	.63
Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_3	.43
Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_4	.79
Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_5	.55
Science_Grade_10_Task_2.10_Concepts_of_Life_Science_Item_6	.51
Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_1	.62
Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_2	.55
Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_3	.61
Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_4	.59
Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_5	.69
Science_Grade_10_Task_3.10_Concepts_of_Earth_Science_Item_6	.24
Science_Grade_10_Task_4.10_Science_and_Technology_Item_1	.65
Science_Grade_10_Task_4.10_Science_and_Technology_Item_2	.50
Science_Grade_10_Task_4.10_Science_and_Technology_Item_3	.65
Science_Grade_10_Task_4.10_Science_and_Technology_Item_4	.71
Science_Grade_10_Task_4.10_Science_and_Technology_Item_5	.82
Science_Grade_10_Task_4.10_Science_and_Technology_Item_6	.82

CHAPTER 9: ANNUAL MEASURABLE OBJECTIVES

Quality assurance is applied to all Annual Measurable Objective (AMO) 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 AMO 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 AMO data file as well as the *Individual Student Reports* (ISRs) generated from the final AMO data file by three additional quality assurance evaluations.

Standard

The first quality assurance evaluation reviewed 14 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 12 reviews, as each performance level was reviewed at each grade level administered; however, two additional student records pulled were ELOS records). Student records were pulled from the AMO file and compared to the original data file.

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 in science.
3) Verified raw scores in science.
4) Verified all AMO performance level assignments.

Individual Student Reports

The second quality assurance evaluation reviewed 14 ISRs, selected to represent all possible permutations of the ISR forms as well as two ELOS records, to ensure that there was 100% match between the final AMO 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 AKAA received ISRs. All students were included.

Confirmatory Quality Assurance Review

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

1) Verified that the reasons not administered are coded appropriately in the final AMO data file.
2) Verified that science scores 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.

Annual Measurable Objectives Report Overview

Annual Measurable Objective (AMOs) 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 science 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.

Science Annual Measurable Objectives

Science Grade 4
In grade 4, approximately 51% of students received a score of 3 (proficient), and 18% of students received a score of 4 (advanced) equaling a total of 69% of all students achieving proficiency. The total percentage of students receiving scores of 2 (below proficient) or 1 (far below proficient) was 31%.

Science Grade 8
In grade 8, approximately 50.6% of students received a score of 3 (proficient), and 18.2% of students received a score of 4 (advanced) equaling a total of 68.8% of all students achieving proficiency. The total percentage of students receiving scores of 2 (below proficient) or 1 (far below proficient) was 31.2%.

Science Grade 10
In grade 10, approximately 55.6% of students received a score of 3 (proficient), and 6.2% of students received a score of 4 (advanced) equaling a total of 61.8% of all students achieving proficiency. The total percentage of students receiving scores of 2 (below proficient) or 1 (far below proficient) was 38.3%.

Appendix 7.1 Science Statistics

CHAPTER 10: PROGRAM IMPROVEMENT

Program Evaluation

The AKAA 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 assessment, and review HelpDesk calls for areas requiring additional training.

Summary of Training Evaluations

Mentors evaluated the October 2014 Annual Mentor Training event.

In general, Mentors believed they were equipped to lead training in their districts on the new DLM assessment system (average score 3.32 of a total possible of 4). Mentors were less confident that their districts were prepared to implement the new Alaska standards and DLM Essential Elements (average 3.14). Comments indicated that the Mentors were concerned about the amount of work still required to implement the new assessments but appreciated the time to work through the online training at the training site.

Only one Mentor completed the online survey evaluation of annual mentor training in January. Several reasons can account for this:

- This was the first time the training evaluation was offered as an online survey; previous evaluations were provided as paper forms distributed and collected on the last day of training

- The Dynamic Learning Maps trainers also asked for training evaluations. Mentors may have believed there was only one evaluation, and the DRA survey was missed

- Participants were informed during the training that DRA was not going to be part of the AKAA next year, so they may have perceived diminished utility in providing feedback for DRA.

Appendix 10.1a October Training Evaluation Summary

Appendix 10.1b January Training Evaluation Form

Summary of Consequential Survey

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

generally positive regarding the social consequences of the AKAA, and very supportive of the role that EED and DRA have played in providing technical assistance and support. Assessors associate no negative consequences with the AKAA, overall. Complete summaries of the quantitative and qualitative results from the survey are provided in Appendix 10.2.

Training and Qualifications

The educators who work with Alaska's most significantly impaired students are experienced and well trained. Responses were received from 103 participants. Qualified Assessors (QAs) made up 77.6% of respondents, Qualified Trainers (QTs) made up 18.9% of the respondents, and 3.5% were Administrators. All respondents had at least a Bachelor's degree, while seventy-one percent of the respondents held Master's degrees and one respondent had a Doctoral degree. Ninety-four percent of the respondents held special education licenses. The majority of respondents administered the AKAA this year, at 94%. The respondents' average years teaching experience was 13.4 years, with an average of 10.2 years teaching SWSCDs. Sixty percent of the respondents administered the Science Alternate Assessment in 2015.

Accessibility and Impact

The majority of respondents agreed or strongly agreed with the statement, "The Alternate Assessment is accessible to my students" and that their students "are improving their academic skills."

Appendix 10.2 Consequential Validity Report

Summary of Help Desk Queries

During the 2015 testing window, DRA's HelpDesk operator (Sevrina Tindal) answered 41 inquiries from Alaska Assessors of the Alternate Assessment. Their queries represented 31 topics.

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 ("AKAA_2015_Helpdesk_Log_VF.xlsx") is located on the Secure Transfer Server. The following table highlights the three most common issues:

Number	Issue Summary
11	Account maintenance issues: Forgotten passwords, new emails
11	Questions specific to DLM assessments and training
4	Questions related to disposal of testing materials

Appendix 10.3 2015 HelpDesk Log

Recommendations for Future Consideration

Technological Improvements

The Alaska Education and Early Development department has elected to contract with a different testing company for the Science Alternate Assessment for the 2015-2016 testing year. Therefore, DRA has no recommendations for improving the current system.

Recommendations for Training

During the 2014-2015 school year, Assessors participated in one or more training venues, including online, face-to-face, and webinars.

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.

In addition, Assessors participated in training related to the Dynamic Learning Maps system for assessing students with significant cognitive disabilities. Assessors could access the DLM training by following instructions detailed on the DRA training site, or could access training directly through DLM's website.

Training Recommendations from Consequential Validity Survey Responses

There were few recommendations from the consequential validity survey that can be applied to the new DLM assessment system, as the assessment approach will be entirely new. The field is both hopeful that the new DLM assessment will effect expected improvements, but is also full of trepidation regarding how much new information must be learned. There are two recommendations from the report that might still be addressed:

1. Incorporate time-saving procedures into the DLM assessment to the degree feasible. A major concern of all QAs and QTs for the AKAA is efficiency in the test administration process. This concern is particularly relevant to the ongoing, formative components of the DLM assessment approach.
2. Effect a discussion regarding how to address the participation of students with severely limited communication and/or medical complications, who are difficult to include in alternate assessment programs.

Training Recommendations from HelpDesk Questions

The category tied with the greatest number of queries (11) related to confusion over regarding the DLM training and testing system. As this was the first year of the DLM assessment system, and Assessors were participating in two separate testing systems, this confusion was expected. Because DRA will not be involved in the AKAA next year, and teachers will become skilled in the new testing system, this is a moot issue.