
# Oregon Department of Education

## 2010–2011 Technical Report

Oregon's Alternate Assessment System

Reliability and Validity

Last updated on August 2009

Oregon's Alternate Assessment System Technical Report: Reliability and Validity



Oregon Department of Education Office of Assessment and Information Services 255 Capitol Street NE Salem, OR 97310 (503) 378-3600 http://www.ode.state.or.us/

*This technical report is one of a series that describes the development of Oregon's Statewide Assessment System. The complete set of volumes provides comprehensive documentation of the development, procedures, technical adequacy, and results of the system.*

**OVERVIEW** ..... **5**
PURPOSE OF OREGON'S ALTERNATE ASSESSMENT ..... 5
INTRODUCTION TO TECHNICAL ADEQUACY ..... 5
APPENDICES ..... 8
**OREGON EXTENDED ASSESSMENT TRAINING 2010-11** ..... **9**
GUIDELINES FOR THE PROVISION OF SUPPORTS ON THE OREGON EXTENDED ASSESSMENTS ..... 12
**ASSESSMENT CONSIDERATIONS** ..... **13**
CONSIDER GENERAL ASSESSMENT WITH OR WITHOUT ACCOMMODATIONS IF ..... 13
CONSIDER STANDARD ADMINISTRATION OF THE EXTENDED ASSESSMENT IF ..... 13
CONSIDER SCAFFOLDED ADMINISTRATION OF EXTENDED ASSESSMENT IF ..... 15
EXTENDED ASSESSMENT ONLINE-TRAINING (WEB-TRAINING) ACCESS INSTRUCTIONS AND QUALIFYING ACTIVITIES ..... 15
**TRAINING AND TEST APPENDICES** ..... **17**
APPENDIX 1.1 ..... 17
APPENDIX 1.2 ..... 17
APPENDIX 1.3 ..... 17
**TEST PREPARATION APPENDICES** ..... **20**
APPENDIX 1.4A ..... 20
APPENDIX 1.4B ..... 20
APPENDIX 1.5 ..... 20
APPENDIX 1.6 ..... 20
APPENDIX 1.7 ..... 20
**PROFICIENCY IN TRAINING RESULTS** ..... **22**
ALL SUBJECTS ..... 22
READING ..... 23
MATH ..... 29
WRITING ..... 35
SCIENCE ..... 41
REFRESHER ..... 47
ADMINISTRATION ..... 54
**DEMOGRAPHICS** ..... **61**
READING ..... 61
WRITING ..... 61
MATH ..... 62
**RELIABILITY** ..... **63**
READING ..... 63
WRITING ..... 65
MATH ..... 66
SCIENCE ..... 70
**DESCRIPTIVE STATISTICS** ..... **73**
READING: ELEMENTARY ..... 73
READING: MIDDLE ..... 74
READING: HIGH ..... 74
WRITING: GRADE 4 ..... 75

WRITING: GRADE 7 ..... 77
WRITING: GRADE 11 ..... 77
MATH: GRADE 3 ..... 78
MATH: GRADE 4 ..... 79
MATH: GRADE 5 ..... 79
MATH: GRADE 6 ..... 80
MATH: GRADE 7 ..... 81
MATH: GRADE 8 ..... 81
MATH: GRADE 11 ..... 82
SCIENCE: GRADE 5 ..... 82
SCIENCE: GRADE 8 ..... 83
SCIENCE: GRADE 11 ..... 84

**ANALYSES WITHIN AND ACROSS SUBJECT AREAS ..... 85**
OVERVIEW ..... 85
CORRELATIONAL ANALYSES ..... 85
REGRESSION METHODS ..... 86

**MODEL 1 RESULTS: PRE-REQ ON SCALE SCORES ..... 90**
READING ..... 90
WRITING ..... 91
MATH ..... 91
SCIENCE ..... 92

**MODEL 2 RESULTS (SIMULTANEOUS): ADMIN TYPE AND PRE-REQ ON SCALE SCORES ..... 93**
READING ..... 93
WRITING ..... 95
MATH ..... 95
SCIENCE ..... 97

**MODEL 3 RESULTS (SEQUENTIAL): DIS, ADMIN, & RACE/ETHNICITY ON PRE-REQ ..... 99**
READING ..... 99
WRITING ..... 100
MATH ..... 101
SCIENCE ..... 102

**MODEL 4 RESULTS (SEQUENTIAL): DIS, ADMIN, & RACE/ETHNICITY ON SCALE SCORE ..... 104**
READING ..... 104
WRITING ..... 105
MATH ..... 106
SCIENCE ..... 107

**MODEL 5 RESULTS: CONTENT TASK TOTALS ON SCALE SCORE ..... 109**
READING ..... 109
WRITING ..... 109
MATH ..... 110
SCIENCE ..... 116

# Overview

This volume provides complete documentation of the alternate assessment in Oregon, its design and development, the technical characteristics of the instruments, and its use and impact in providing proficiency data on grade level state standards as part of the mandates from No Child Left Behind (NCLB).

## Purpose of Oregon's Alternate Assessment

In this technical report, we present data to support the claim that Oregon's Alternate Assessment provides the state technically adequate student performance data to ascertain proficiency on grade level state content standards for students with significant disabilities. The alternate assessments based on alternate achievement standards are aligned with grade level academic content; generate reliable outcomes at the item, task, and test level; include all students; have a cogent internal structure; and fit within a network of relations within and across various dimensions of content related to and relevant for making proficiency decisions.

## Introduction to Technical Adequacy

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) reliability, (e) validity, and (f) other dimensions of technical quality. This document addresses the latter four requirements 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 and response process, then reliability, and finally address 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*

The *content related evidence section* provides information on test content, training, and proficiency on test administration. In particular, the content reflects '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. Training was conducted in both regional sessions and through web-based modules.

Evidence of content coverage is concerned with judgments about "the adequacy with which the test content represents the content domain" (AERA et al., 1999, p. 11)[7]. As a whole, the test is comprised of sets of items that sample student performance on the intended domains. The expectation is that the items cover the full range of intended domains, with a sufficient number of items so that scores credibly represent student knowledge and skills in those areas. Without a sufficient number of items, the potential exists for a validity threat due to construct under-representation (Messick, 1989)[4].

Our foundation of validity evidence from content coverage comes in the form of test blueprints or test specifications. Among other things, the *Standards* (AERA et al., 1999)[7] suggest specifications should "define the content of the test, the number of items on the test, and the formats of those items" (Standard 3.3, p. 43).[5]

All items and tasks are linked to grade level standards and a prototype was developed using principles of universal design with traditional item writing techniques. The most important component in these initial steps addressed language complexity and access to students using both receptive, as well as expressive, communication. Additionally, both breadth and depth were addressed. We developed two forms of each grade level test, a standard and a scaffold version. The scaffold administration utilizes a more accommodated approach that allows for students with very limited attentional resources to access the same test content as their peers who participate in the standard version. The test is designed to be comparable across multiple disabilities, with prerequisite skills and test type accounting form most of the variance. Any differences between the assessments are thus deemed to be construct-irrelevant (see Method 2 in chapter 4). In each task, we generally increased the depth of knowledge from the first to the last item.

We developed the test iteratively by developing items and tasks, piloting them, reviewing them, and editing successive drafts. We used existing panels of teachers who have worked with the Oregon Department of Education in various advising roles on testing content in general education, using the same processes and criteria. While the internal reviews of content were initially conducted within Behavioral Research and Teaching, after the initial draft of prototype items, all reviews involved content experts with K-12 classroom experience. The first level review was to ensure universal design and incorporated two experts to represent the blind and deaf communities. Finally, subsequent reviews were conducted to ensure appropriate administration and scoring, all of which was completed as part of training.

The *response process* section presents outcome data on the manner in which students took the tests. First we address the training in administration. Because we had designed a flexible performance assessment with various options for teachers to use in testing students, it was

---

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

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

imperative to have a work force fully informed. For the 2010 administration of the Oregon Extended Assessment we updated the training structure that included qualified mentor/trainers and a web-based training and proficiency system that required all users to pass a qualifying test. We report high levels of knowledge and proficiency levels on the qualifying test.

The *reliability section* presents three types of analyses: (a) internal consistency for each task in each subject area for every grade level, (b) inter-item correlations, and (c) reliability from administration. The test has high reliability in every task and subject area.

The *criterion-related evidence section* documents how well the test fits within a network of relationships. Because we had designed an assessment that first documented the student's access skill (pre-requisite skill) to assist teachers in presenting the content items, we also describe the options for participation using pre-requisite skills assessments to allow teachers use of various levels of support. Pre-requisite skills were assessed to provide the necessary supports for appropriate test administration (with four levels: full physical support, partial physical support, prompted support, and no support). Content prompts were designed to document students' skill and knowledge on grade level academic content standards. We also designed two test administration types that Individualized Educational Program (IEP) teams could choose to use: (a) standard or (b) scaffold. Both types addressed exactly the same content and only differed in the amount of scaffold they provided to access the target skill (content prompt).

Perhaps the best model for understanding criterion-related evidence comes from Campbell and Fiske (1959)[6] in their description of the multi-trait, multi-method analysis. N. B. we translate the term 'trait' to mean 'skill']. In this process (several) different traits are measured using (several) different methods to provide a correlation matrix that should reflect specific patterns supportive of the claim being made (that is, provide positive validation evidence). Sometimes, these various measures are of the same or similar skills, abilities, or traits, and other times, they are of different skills, abilities, or traits. We present data that quite consistently reflects higher relations among tasks **within** an academic subject than **between** academic subjects. We also present data in which performance on content prompts is totaled within categories of disability, expecting relations that would reflect appropriate differences (see Tindal, McDonald, Tedesco, Glasgow, Almond, Crawford, & Hollenbeck, 2003).[7]

Given the content-related evidence that we present related to test development, administration, and scoring, the response processes related to the levels of independence, the reliability information reflected by adequate coefficients for tasks and tests, and finally, the relation of tasks within and across subject areas (providing criterion-related evidence), we conclude that the alternate assessment judged against alternate achievement standards allows valid inferences to be made on state accountability proficiency standards.

---

[6] Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait, multi-method matrix. In W. A. Mehrens & R. L. Ebel (Eds.), *Principles of educational and psychological measurement: A book of selected readings* (pp 273-302). Chicago, IL: Rand McNally & Company.

[7] Tindal, G., McDonald, Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children*, 69(4), 481-494.

## Appendices

| Topic | File Name |
|---|---|
| Slides for training new qualified assessors, new qualified trainers, and returning assessors | App1.1_QA_QT_ReturningTraining.pdf |
| General Administration and Scoring Manual | App1.2_AssessmentAdministrationManual_2010-11v3 |
| Sample test items in reading, writing, mathematics, and science | App1.3_RWMSsampleitems |
| | |
| Webinar training for mathematics alignment study | App1.4a_OR_ExMath_AlignmentWebinar_01-12-2011 |
| Results from mathematics alignment study | App1.4b_Oregon Ex Math Alignment Study_2_16_2011_v2 |
| | |
| Technical specifications for science test | App1.5_SciTstSpcs1112_v3 |
| Results from science alignment study | App1.6_OregonExAssessScienceAlignmentStudy_v4 |
| Results from science standard setting | App1.7_OregonExtendedAssessmentScienceStandardSettingReport |
| | |
| Demographics for participants in 2010-2011 alternate assessment | AppA_Dems |
| Reliability of tasks and tests in reading, writing, mathematics and science for all grade levels | AppB_Rel_v2 |
| Descriptive statistics for all tasks in reading, writing, mathematics and science for all grade levels | AppC_Dscrpt |
| Correlations across subject areas | AppD_Corr |
| Regression model using pre-requisite skills as the outcome | AppE_Model1PreReq |
| Regression model using pre-requisite skills and type of administration as the outcome | AppF_Model2Adm&PreReq |
| Regression model using disability to predict pre-requisite skills | AppG_Model3DisOnPreReq |
| Regression model using disability to predict RIT scores | AppH_Model4DisOnRIT |
| Regression model using tasks to predict RIT scores | AppI_Model5TskOnRIT |

## Oregon Extended Assessment Training 2010-11

The Oregon Department of Education (ODE) provided statewide training for three applications in the 2010-11 school year. First, they provided training for new Qualified Assessors (QAs) via webinar and through the Qualified Trainer structure. Second, they provided direct training for new Qualified Trainers (QTs) via five regional trainings held in Hillsboro, Salem, Redmond, Medford, and Pendleton. Third, returning Assessors were trained via webinar. The schedule for the regional trainings, as well as relevant training information, is provided below:

| Date | Who/Team | Location |
|---|---|---|
| | | |
| Nov. 3rd | Team: DC, BL, SJ<br>Contact:<br>Kerri Smith<br>kmsmith@nwresd.k12.or.us<br>503-614-1428<br>Room Reserved –6/29 | NWRESD-<br>5825 NE Ray Cir<br>Hillsboro, OR 97124-6436 |
| Nov. 10th | Team: DC, BL, JT and/or SJ<br>Contact:<br>Tom Beach<br>Tom.beach@wesd.org<br>503-588-5330 | WESD-<br>2611 Pringle Rd SE<br>Salem, OR 97302-1533 |
| Nov 16th | Team: DC, JT<br>Contact:<br>Catherine Kelly<br>catherine.kelly@hdesd.org<br>541-693-5702 | HDESD-<br>145 SE Salmon Ave<br>Ste A<br>Redmond, OR 97756-8427 |
| Nov 16th | Team: BL, SJ<br>Contact:<br>Marian Gerstmar<br>marian_gerstmar@soesd.k12.or.us<br>541-776-8555 | SOESD-<br>101 Grape St.<br>Medford, OR 97501-2793 |
| Nov. 18th | Team: BL, SJ<br><br>Contact:<br>Mary Apple<br>mary.apple@umesd.k12.or.us<br>541-276-6616 | UMESD-<br>2001 SW Nye Ave<br>Pendleton, OR 97801-4416 |

## NEW Qualified Assessor (QA) and Qualified Trainer (QT) Extended Assessment Training for 2010 - 2011: What Districts Need to know

QUALIFIED ASSESSOR

*When will ASSESSOR training sessions occur?*
November - February (as needed)

*Where will Assessor trainings occur?*
In districts and ESDs across the state

*How many Qualified Assessors will need to be trained in my district?*
The number/need will vary by district. Any individuals who will or may administer Extended Assessments in the current school year should be trained.

*How will Qualified Assessors be selected?*
Qualified Assessors will be selected at the district/local level by administrators and qualified educators (e.g. DTCs and individuals knowledgeable about Extended Assessments). Previous experience with alternate assessments is not required to become a QA.

*How long will the training take?*
Approximately 4 – 8 hours depending on need--as determined by the Qualified Trainer.

*What assessments will the individuals be trained on?*
Extended Assessments (at grade level(s) as determined by need). Administration of new field test items (science only this year).

*How will the trainings be conducted?*
Locally as arranged and conducted by Qualified Trainer(s) within state and district parameters.

*Who will pay any costs associated with QT training?*
As in previous years, districts and ESDs will be allocated funds based on their SECC that are to designated for activities related to the assessment of students with disabilities.

QUALIFIED TRAINERS (QT)

*When will TRAINER training sessions occur?*
November

*Where will new trainer trainings occur?*
Five live trainings in five zones across the state

*How many Qualified Trainers will be trained in my district?*
ODE's goal is to retain and recruit approximately 200 - 300 individuals from across the state according to child count/district size. Sufficient trainers should be trained to ensure ongoing capacity in your district.

*What qualifications are necessary to attend the training?*
Individuals who were Qualified Assessors in 2009-2010 and with permission from their district may train to become Qualified Trainers in 2010-11.

*How long will the live trainer training sessions take?*
Approximately five hours of training

*What assessments will the individuals be trained on?*
Extended Assessment (Elementary, Middle, and High school levels) Administration of new field test items for science only.

*How will the trainings be conducted?*
State-level trainers will conduct trainings at five sessions held at regional sites.

*Who will pay any costs associated with QT training?*
As in previous years, districts and ESDs will be allocated funds based on their SECC that are to designated for activities related to the assessment of students with disabilities.

## RETURNING Qualified Trainer Extended Assessment "Refresher" Sessions for 2010 - 2011: What Districts Need to know

*What is the purpose and focus of the "Refresher" sessions?*
The purpose and focus of the "Refresher" session is to provide a venue for current, returning Qualified Trainers only to receive information related to the Extended Assessment (i.e., administration, data entry, science field test, training and proficiency site, and QT role and responsibilities).

*When will Trainer "Refresher" sessions occur?*
Three opportunities to participate via webinar (WebEx) will be offered: October 19th (9a-Noon), January 5th (1p-4p), and January 26th (9a-Noon).

*What qualifications are necessary to attend a "Refresher" webinar?*
Individuals who were Qualified Trainers the prior school year and are planning to retain their status for the current school year.

*Who will pay any costs associated with attending a "Refresher" webinar?*
As in previous years, districts and ESDs will be allocated funds based on their SECC that are to designated for activities related to the assessment of students with disabilities.

*Registration: To register for a webinar please contact Linda O'Hallaran at* Linda.OHallaran@state.or.us. Once we receive your contact information we will send you the information necessary to access the webinar for which you have registered. *Access to the internet and a phone will be necessary to participate*. Information on Extended Assessments and Trainings are posted at http://www.ode.state.or.us/search/page/?id=2689.

## Qualified Trainer Action Items: Building Capacity as a Qualified Trainer

1. Identify yourself to your District office as a Qualified Trainer.
2. Inform your District office that individuals will be calling to determine names of Qualified

Trainers and that you would like your name made available (funding will be provided to support additional training toward capacity building in your local areas, how much funding varies by district. Districts may provide other funds to support this training)
3. Work with your schools or districts to determine…
   a. Adequate locations to provide training
   b. Support for printing materials
   c. Any other supports that may/will be provided for the meeting
   d. How many Qualified Assessors the district will need
4. Create a flyer or determine some other method of advertising your Qualification and your intent/availability to provide trainings. In the flyer, identify:
   a. dates
   b. times
   c. length of training and the level of training

It is advisable to separate training so that those individuals who are familiar with the assessments receive a separate/shorter training than those receiving training for the first time
If you are training novices, you will need to provide training on the:
   a. Format
   b. Administration
   c. Selection of Science Field test materials
   d. Scoring
   e. Data entry (including Science Field test)
   f. Interpretation of responses
   g. Overall process
   h. Locations of websites (Training, Data Entry, Resource)

If you are training individuals who received some training last year you will need to train on:
   a. Changes (including selection of Science Field Test materials), Scoring prerequisites, Data entry (including Science Field test), Providing support during OR the Extended Assessments, Website
   b. Direct previously Qualified Assessors (qualified in 09-10) to re-qualify independently online through the Extended Assessment training site.

Your Contact for general questions related to your district's needs is your district or ESD Special Education Director unless otherwise indicated.

**Guidelines for the Provision of Supports on the Oregon Extended Assessments**
Supports provided during the administration of the Prerequisite Skills items serve a different purpose than supports provided during the administration of the Content Prompts. During the Prerequisite Skills items, a student may be supported to success, however during the administration of the Content Prompts a student is supported to access only.

**Full Physical Support**
Provided for students who routinely need full physical supports to participate in instruction. Full physical support is not to be given to a student who does not receive full physical support in the instructional environment. Full physical support is reserved for those students with significant mobility impairments who, as a result, rely on these supports routinely)
- Assisting with positioning toward response options
- Assisting with positioning toward correct response options
- Positioning student's hand on correct response following progressive movement along the continuum of supports from full independence (as needed)

- Based on prolonged hesitation or an indication of student uncertainty, assessor provides any (or a combination) of the following:
  - Moving student to materials
  - Positioning student in a responding position in relation to the materials
  - Orienting student to the appropriate response options in the materials
  - Moving student's hand over a series of response options in the materials

**Partial Physical Support**

- Gentle movement of the student's hand (prompting) toward the materials
- Physical repositioning if student selects a non-response option

Based on prolonged hesitation or an indication of student uncertainty, assessor provides any (or a combination) of the following:

- Touch student to direct his/her attention toward the appropriate materials
- Touching student to determine/obtain attention

**Visual, Verbal, or Gestural Support**

- Visual: Physical adjustment of the materials so that they are in a optimal visual location for the student's needs.
- Verbal: Additional verbal directions about the item
- Gestural: Specific gesturing toward the materials to indicate the intent of the item, or pointing to the correct item
- Based on prolonged hesitation or an indication of student uncertainty, assessor provides any (or a combination) of the following:
  - Visual: Maintaining optimal visual placement of assessment materials for student (i.e. moving materials to ensure they remain within student gaze)
  - Verbal: Rephrasing *process* directions:

  "You are choosing from these three" "You are putting these in order" "You are telling me yes or no"
  - Gestural: Specific gesturing to the student to get student's attention

**Full independence**

Student needs no supports to perform the item successfully.
Student needs no supports to gain access to the structure of the item or the associated materials.

## Assessment Considerations

**Consider General Assessment with or without accommodations if…**

*Student:*

- Performs at or around grade level
- Has academic difficulties that primarily surround reading but may be average or close to average in other subject areas
- Has academic difficulties in areas other than reading that are "mild to moderate" and can typically be addressed by using simplified language
- Is reading within two to three grades of his or her enrolled level

*Instruction:*

- Is primarily general curriculum instruction (but may also use a specialized curriculum in some areas)

*Some Judgment variables:*

- What assessment did he take last year?
- How is his attention?
- What types of behaviors should be considered?

**Consider Standard Administration of the Extended Assessment if…**

*Student:*

- Performs well below grade level
- Is significantly below grade level in reading
- Has academic difficulties that are generalized (to all subject areas) and are significant
- Benefits from specialized individual supports

*Instruction:*

- Is primarily from a specialized curriculum
- From general curriculum must be significantly reduced in breadth, depth, and complexity

*Some Judgment variables:*

- What assessment did he take last year?
- How is his attention?
- What types of behaviors should be considered?
- Previous relevant experiences

**Consider Scaffolded Administration of Extended Assessment if…**

*Student:*
Performance is significantly impacted by the disability, the student does not read, has academic, mobility, and receptive and expressive language difficulties that are generalized and significant, and/or relies on individual and significant supports to access reduced content materials.

*Instruction:*
Is from a specialized curriculum and has functional components and/or includes academic goals that are significantly reduced in depth, breadth, and complexity from grade level content.

*Some Judgment variables:*
Is the student able to interact with instructional material in a way that provides meaningful feedback?

**Extended Assessment online-training (Web-Training) Access Instructions and Qualifying Activities**

1) What do I do if I was a Qualified Assessor (QA) for the Extended Assessments last year (2009- 10) and would like to upgrade/refresh online to assess students again this year? If you successfully qualified as a Qualified Assessor (QA) during the 2009-10 school year and you intend to remain a QA during 2010-11 you must refresh your qualifications to retain your status. For the 2010-11 school year you will need to do the following to refresh (or reactivate) your qualification status online.

a) Go to* the Oregon Training and Proficiency website: http://or.k12test.com (*if your email address has changed or you don't remember your password see "Technical Difficulties" box on this page) 2) Review the site for any changes throughout and read the "Updates 2010-11" part of the Training section

b) Take the Refresher Proficiency test, a 25 question test and pass with a score of at least 80% 4) Once you pass the Refresher Proficiency you will automatically be up upgraded to Qualified Assessor status. A "Materials" section will now appear on your account, which provides practice tests for your use.

2) What do I do if I was a Qualified Trainer (QT) for the Extended Assessments last year and would like to upgrade/refresh online to either train assessors or to assess students again this year? If you successfully qualified as a Qualified Trainer (QT) during the 2009-10 school year and you intend to remain a QT and/or assess students with the Extended Assessment this year (2010-11), you must refresh your qualifications to retain your status. For the 2010-11 school year you will need to do the following to refresh (or reactivate) your qualification status online:

a) Participate in a "Refresher Session" webinar sponsored by the Oregon Department of Education (see http://www.ode.state.or.us/search/results/?id=178 to register).

b) Go to* the Oregon Training and Proficiency website: http://or.k12test.com (*if your email address has changed or you don't remember your password see "Technical Difficulties" box on this page).

c) Read the "Updates 2010-11" part of the Training section. 4) Take the Refresher Proficiency test, a 25-question test and pass with a score of at least 80%. 5) Once you pass the Refresher Proficiency you will automatically be upgraded to Qualified Trainer status. A "Materials" section will now appear on your account, which provides materials

for training New Qualified Assessors.

3) What do I do if I began the Extended Assessment training process last year (either QT or QA), but did not complete it? If you attended any live training last year (either QT or QA trainings) but did not complete the online training and proficiency process, your registration information was not retained in Oregon's Extended Assessment web-based training system. In Oregon, only Qualified Assessors (QA) or Qualified Trainers (QT) are allowed to administer Extended Assessments to students. If you wish to become a Qualified Assessor you must do the following:

a) Attend a live training that is conducted by a Qualified Trainer in your district/region during this school year.

b) After you complete the live training, log onto the training and proficiency website: http://or.k12test.com

c) Complete the training section of the website.

d) Pass the 5 proficiency tests with a score of at least 80%.

e) Once you have passed the proficiency tests your status will be automatically upgraded to a Qualified Assessor.

*Technical Difficulties

If your email address has changed: 1) Go to the website, http://or.k12test.com, and log in using your old email address and password 2) Go to the "Account" section 3) To the right of the email address click on the blue link that reads "Update" 4) Enter your new email address in the box and click on "Send Confirmation" 5) Go to your new email account and click on the email that was just sent to you with a subject title of "Oregon Extended Registration Confirmation Link" 6) Within the body of that email, click on the Confirmation Link, which will take you back to the training and proficiency website, and your email address has been changed. If you don't remember your password: 1) Go to the website, http://or.k12test.com, and click on the blue words "Reset Password" 2) Enter your email address that you used last year (or the new email address you changed to this year), and click on "Send Email Confirmation" 3) Go to your email account and click on the email that was just sent to you with a subject title of "Oregon Extended Password Reset Confirmation" 4) A website page will be generated and your new password is on the first line in green. Copy the password and click on the blue "login" word. 5) Enter your email address and the new password and login. 6) Go to the "Account" section of the website once logged in, and change your password if you wish.

## Training and Test Appendices

**Appendix 1.1**
Appendix 1.1 is the PowerPoint training that was used by ODE and BRT trainers to train New and Returning Qualified Trainers (QTs) in five regional trainings in November 2010. QTs also used the package to train New Qualified Assessors for the 2010-11 school year. The training provides participants with the information needed to pass proficiency tests as part of the requirements to become a Qualified Assessor for the Oregon Extended Assessments and was delivered by QTs throughout the state. The training package addresses the following topics:

- Eligibility – which students take AA-AAS?
- Standard Administration/Scaffold Administration?
- 2011 Test Window
- Student Confidentiality & Test Security
- Test Administration (Physical & Logistic)
- Scoring & Data Entry
- Reports & Sharing Results with Parents
- 2011 Science Field Test Plan
- Math, Reading, and Writing Field Test Information
- What's new in 2010-11
- Resources

**Appendix 1.2**
Appendix 1.2 is ODE's General Administration and Scoring Manual for 2010-11. The manual establishes ODE's expectations regarding the test window, utilizing the OR Extended website, and informing parents. It also provides the following information for stakeholders, including educators and parents:

- Overview of the Extended Assessments
- Assessing a Student
- Scoring
- Decision Making
- Information for Teachers.

The manual provides three appendices that provide guidance regarding the provision of supports, parent questions and answers, and a glossary.

**Appendix 1.3**
Appendix 1.3 provides stakeholders with visual representation of the structure of the Oregon Extended Assessment. Sample tasks/items are conveyed, including both Prerequisite and Content Prompts. There are standard and scaffold administration tasks represented in reading, writing, math, and science. The appendix shows what a Qualified Assessor would be viewing during test administration (Scoring/teacher's Protocol) as well as what the student would be viewing as the QA asks the test questions (Student Materials). Stakeholders can see the structure of each task/item, as well as how the items are scored. They can also gather an idea about the types of formats that are used for answer choices that are included within the Student Materials documents.

| Topic | File Name |
|---|---|
| Slides for training new qualified assessors, qualified | App1.1_QA_QTReturningTraining.pdf |

| | |
|---|---|
| trainers, and returning qualified trainers | |
| General Administration and Scoring Manual | App1.2_AssessmentAdministrationManual_2010-11v3 |
| Sample test items in reading, writing, mathematics, and science | App1.3_RWMSsampleitems |

PLEASE INDICATE HOST LOCATION: _____

EXTENDED ASSESSMENT REGIONAL TRAINING EVALUATION

**PLEASE complete this evaluation

| 9:15-12:00 | Overview: Information, Guidance, & Updates | | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| The materials for this section of the training were appropriate, relevant, and useful. | | | | | |
| The presenter(s) was knowledgeable, organized, and clear. | | | | | |
| I felt comfortable with the resources presented, both online and those provided in hardcopy. | | | | | |
| After attending this section, I am better able to explain this topic to colleagues and constituents. | | | | | |
| Overall, I rate this section as useful. | | | | | |
| Please share suggestions or comments: | | | | | |
| | | | | | |
| | | | | | |

| 12:30-2:00 | Using and Navigating the Web-based system | | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| The materials for this section of the training were appropriate, relevant, and useful. | | | | | |
| The presenter(s) was knowledgeable, organized, and clear. | | | | | |
| I felt comfortable with the resources presented, both online and those provided in hardcopy. | | | | | |
| After attending this section, I am better able to explain this topic to colleagues and constituents. | | | | | |
| Overall, I rate this section as useful. | | | | | |
| Please share suggestions or comments: | | | | | |

1 = strongly disagree 2 = disagree 3 = agree 4 = strongly agree

## Test Preparation Appendices

**Appendix 1.4a**
Appendix 1.4a is the training package that was used to train reviewers for the January 12, 2011 Oregon Extended Math Alignment Study. The training was conducted via webinar, utilizing BRT's Distributed Item Review (DIR) secure website for file transfers. The training elaborates the item development process, including how ODE and BRT approach reductions in terms of depth, breadth, and complexity. A four-point alignment scale is established (0= no link to the standard, up to 3=direct link to the standard), which elaborates the coding system used by participants to determine the level of linkage between the item and the standard. Participants are guided through multiple examples to increase their judgment consistency. Participants were asked to rate the linkage of each item and provide comments for items with lower than expected levels of linkage.

**Appendix 1.4b**
Appendix 1.4b is the summative report of the *Oregon Alternate Assessment 2011 Alignment Study in Mathematics*, completed by Dr. Lindy Crawford on 2-12-2011. The report is divided into three sections that provide information regarding the professional background of all reviewers, the item review process, and a summary of study outcomes.

**Appendix 1.5**
Appendix 1.5 is the *Oregon Extended Science Test Specifications and Test Blueprint*. The document provides background information regarding test structure, item development, the structure of Oregon's new science standards, the changes that have occurred to the Science Extended Assessment in the 2010-11 school year, and the 2011-12 test blueprint that was developed based upon the input gathered from alignment study and the standard setting procedures.

**Appendix 1.6**
Appendix 1.6 is the May 4, 2011 *Oregon Alternate Assessment 2011 Alignment Study in Science*, completed by Dr. Lindy Crawford on May 4, 2011. The report is organized into three sections, including a description of all reviewers, the item review process, and a summary of outcomes from the study.

**Appendix 1.7**
Appendix 1.7 is *Oregon's Science Alternate Assessment 2011 Standard Setting Evaluation* report. The document includes the agenda for the meeting, the professional background information for all panelists, the process used to develop achievement level descriptors (ALDs), outcomes for each group in Rounds 1, 2, and 3, as well as panelist confidence ratings and input regarding the process of developing ALDs and setting cut scores for the Oregon Extended Science Assessment.

| Topic | File Name |
|---|---|
| Webinar training for mathematics alignment study | App1.4a_OR_ExMath_AlignmentWebinar_01-12-2011 |
| Results from mathematics alignment study | App1.4b_Oregon Ex Math Alignment Study_2_16_2011_v2 |
| | |
| Technical specifications for science test | App1.5_SciTstSpcs1112_V2.doc |
| Results from science alignment study | App1.6_OregonExAssessScienceAlignmentStudy_v4 |
| Results from science standard setting | App1.7_OregonExtendedAssessmentScienceStandardSettingReport |

## Proficiency in Training Results

The tables in this section provide statistical information by subject area regarding the proficiency tests that all Qualified Assessors (QAs) are required to pass in order to administer the Oregon Extended Assessments.

The first table is a summary table for Administration, all four subjects (Reading, Writing, Math, Science), and the Refresher training for Returning QAs. It provides information regarding how many participants took each subject area proficiency test (N), the minimum score possible (Minimum), the maximum score possible (Maximum), the average test score for that subject (Mean), as well as a measure of variance by each subject area, or how close to the mean the average examinee achieved on each proficiency test (Std. Deviation).

The tables thereafter are organized by subject area and follow the same format. The prompt is presented in bold font above the table. Item statistics are elaborated for each prompt within the table. The tables include information regarding the number of New QAs who participated, as well as the number of Returning QAs who participated. These two groups show up in some cases as the 'Missing' data from each other's tables, because they took only the assessment prescribed to them (e.g., New QAs had to take all four subject area and administration proficiency tests; Returning QAs only had to take the Refresher proficiency test). Other data is missing because the person is in the database but did not participate in the test.

There are 20 Administration, Reading, Math, Writing, and Science proficiency test items. There are 25 Refresher proficiency test items. The answers are either incorrect (0) or correct (1). The number of total participants who answered incorrectly/correctly is provided (Frequency). The percentage of total participants (those who participated in the full training and those who took the refresher training) who answered incorrectly/correctly is provided (Percent). The percentage of participants who actually participated in the full or refresher test, as appropriate, is provided (Valid Percent), as is the percentage of participants who answered incorrectly and then correctly (Cumulative Percent). The results reflect appropriate levels of difficulty and represent a broad range of topics germane to the administration and scoring of the Oregon Extended Assessments.

### ALL SUBJECTS

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Administration Total | 389 | 1 | 20 | 17.07 | 2.473 |
| Reading Total | 389 | 1 | 20 | 18.61 | 2.054 |
| Math Total | 384 | 5 | 20 | 19.40 | 1.275 |
| Writing Total | 378 | 11 | 20 | 17.49 | 1.860 |
| Science Total | 375 | 15 | 20 | 19.68 | .650 |
| Refresher Total | 1047 | 1 | 25 | 22.80 | 2.567 |
| Valid N (listwise) | 0 | | | | |

## READING

The reading proficiency assessment evaluates Assessor understanding of the administration and scoring requirements related to the Oregon Extended Assessments in the domain of reading. Prerequisite skills tasks and content tasks are represented in the assessment, as are Scaffold and Standard versions of the Oregon Extended Assessments.

**The score the student should receive for item 3, Which one is a letter?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 21 | 1.2 | 5.4 | 5.4 |
| | 1 | 368 | 20.4 | 94.6 | 100.0 |
| | Total | 389 | 21.5 | 100.0 | |
| Missing | System | 1417 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Show me a letter that makes the mmm sound**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 42 | 2.3 | 10.8 | 10.8 |
| | 1 | 346 | 19.2 | 89.2 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 8, Where is a story?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 20 | 1.1 | 5.2 | 5.2 |
| | 1 | 368 | 20.4 | 94.8 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 9, Where is the title of this story?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 16 | .9 | 4.1 | 4.1 |
| | 1 | 371 | 20.5 | 95.9 | 100.0 |
| | Total | 387 | 21.4 | 100.0 | |
| Missing | System | 1419 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 10, Where is the end of the story?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 10 | .6 | 2.6 | 2.6 |
| | 1 | 378 | 20.9 | 97.4 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, Which week had the lowest average score?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 13 | .7 | 3.4 | 3.4 |
| | 1 | 375 | 20.8 | 96.6 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Which week had the highest average score?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 141 | 7.8 | 36.3 | 36.3 |
| | 1 | 247 | 13.7 | 63.7 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3,**
**Did the average scores go up, go down, or stay the same over the weeks?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 4 | .2 | 1.0 | 1.0 |
| | 1 | 384 | 21.3 | 99.0 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Which school has the most 7th grade students?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 8 | .4 | 2.1 | 2.1 |
| | 1 | 378 | 20.9 | 97.9 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Which school has the fewest 7th grade students?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 140 | 7.8 | 36.3 | 36.3 |
| | 1 | 246 | 13.6 | 63.7 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, Which is another word for enormous?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 6 | .3 | 1.6 | 1.6 |
| | 1 | 380 | 21.0 | 98.4 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Which is another word for peered?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 9 | .5 | 2.3 | 2.3 |
| | 1 | 377 | 20.9 | 97.7 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Which is another word for secure?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 381 | 21.1 | 98.7 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, What does keep an eye on mean?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 9 | .5 | 2.3 | 2.3 |
| | 1 | 377 | 20.9 | 97.7 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, What does over his head mean?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 30 | 1.7 | 7.8 | 7.8 |
| | 1 | 356 | 19.7 | 92.2 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, What is this story mostly about?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 1 | .1 | .3 | .3 |
| | 1 | 385 | 21.3 | 99.7 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Which one grows low to the ground?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 9 | .5 | 2.3 | 2.3 |
| | 1 | 377 | 20.9 | 97.7 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, What is one way a shrub is different from a tree?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 383 | 21.2 | 99.2 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, What can people do to give trees & shrubs shape?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 8 | .4 | 2.1 | 2.1 |
| | 1 | 378 | 20.9 | 97.9 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Which is a place you can find shrubs?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 383 | 21.2 | 99.2 | 100.0 |
| | Total | 386 | 21.4 | 100.0 | |
| Missing | System | 1420 | 78.6 | | |
| Total | | 1806 | 100.0 | | |

## MATH

The math proficiency assessment evaluates Assessor understanding of the administration and scoring requirements related to the Oregon Extended Assessments in the domain of math. Prerequisite skills tasks and content tasks are represented in the assessment, as are Scaffold and Standard versions of the Oregon Extended Assessments.

**The score the student should receive for item 3, Which is the number 3?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 29 | 1.6 | 7.6 | 7.6 |
| | 1 | 355 | 19.7 | 92.4 | 100.0 |
| | Total | 384 | 21.3 | 100.0 | |
| Missing | System | 1422 | 78.7 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Which one is a triangle?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 23 | 1.3 | 6.0 | 6.0 |
| | 1 | 361 | 20.0 | 94.0 | 100.0 |
| | Total | 384 | 21.3 | 100.0 | |
| Missing | System | 1422 | 78.7 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 8, Which circle is the smallest?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 50 | 2.8 | 13.0 | 13.0 |
| | 1 | 334 | 18.5 | 87.0 | 100.0 |
| | Total | 384 | 21.3 | 100.0 | |
| Missing | System | 1422 | 78.7 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 9, Which coin is a penny?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 16 | .9 | 4.2 | 4.2 |
| | 1 | 368 | 20.4 | 95.8 | 100.0 |
| | Total | 384 | 21.3 | 100.0 | |
| Missing | System | 1422 | 78.7 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 10, Which one is a decimal?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 12 | .7 | 3.1 | 3.1 |
| | 1 | 372 | 20.6 | 96.9 | 100.0 |
| | Total | 384 | 21.3 | 100.0 | |
| Missing | System | 1422 | 78.7 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1,**
**Which fraction shows how the pie should be cut to make 2 equal pieces?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 12 | .7 | 3.1 | 3.1 |
| | 1 | 371 | 20.5 | 96.9 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Which fraction is equal to 1?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 378 | 20.9 | 98.7 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3,**
**Which set of fractions, A, B, or C, is in order from smallest to largest?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 10 | .6 | 2.6 | 2.6 |
| | 1 | 373 | 20.7 | 97.4 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Which fraction is equal to 1/2?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 1 | .1 | .3 | .3 |
| | 1 | 382 | 21.2 | 99.7 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Add these fractions.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 4 | .2 | 1.0 | 1.0 |
| | 1 | 379 | 21.0 | 99.0 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, Which number comes after 0.4?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 378 | 20.9 | 98.7 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Which decimal is the same as the fraction 3/10?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 378 | 20.9 | 98.7 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Which number is greater than 1/2?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 378 | 20.9 | 98.7 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, About how much money did Brianne spend in all?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 380 | 21.0 | 99.2 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, How much did he spend in all?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 4 | .2 | 1.0 | 1.0 |
| | 1 | 379 | 21.0 | 99.0 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, When x=4, what does y equal?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 7 | .4 | 1.8 | 1.8 |
| | 1 | 376 | 20.8 | 98.2 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, How many feet does the tree grow each month?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 380 | 21.0 | 99.2 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, What is the y-intercept of this line?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 378 | 20.9 | 98.7 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, When x=2, what does y equal?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 380 | 21.0 | 99.2 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Which pair of equations, A, B, or C, has the same intercept?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 13 | .7 | 3.4 | 3.4 |
| | 1 | 370 | 20.5 | 96.6 | 100.0 |
| | Total | 383 | 21.2 | 100.0 | |
| Missing | System | 1423 | 78.8 | | |
| Total | | 1806 | 100.0 | | |

## WRITING

The math proficiency assessment evaluates Assessor understanding of the administration and scoring requirements related to the Oregon Extended Assessments in the domain of writing. Prerequisite skills tasks and content tasks are represented in the assessment, as are Scaffold and Standard versions of the Oregon Extended Assessments.

**The score the student should receive for item 5, Circle/Point to the word with a question mark?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 95 | 5.3 | 25.1 | 25.1 |
| | 1 | 283 | 15.7 | 74.9 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**For item 6, Which word is underlined? The Assessor makes a slight mistake of**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 44 | 2.4 | 11.6 | 11.6 |
| | 1 | 334 | 18.5 | 88.4 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**For item 7, Trace the word exactly (run) the Assessor points to where the item is in the Student Materials, which is not providing any additional support. The score the student should receive is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 20 | 1.1 | 5.3 | 5.3 |
| | 1 | 358 | 19.8 | 94.7 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 8, Copy the word exactly (two)**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 12 | .7 | 3.2 | 3.2 |
| | 1 | 366 | 20.3 | 96.8 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 9, Write the letter T**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 13 | .7 | 3.4 | 3.4 |
| | 1 | 365 | 20.2 | 96.6 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**After administering these 10 Prerequisite Skills items to the student, there were 5 - 4s, 2 - 3s, 3 - 1s.**

**The Independence for Access score is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 40 | 2.2 | 10.6 | 10.6 |
| | 1 | 338 | 18.7 | 89.4 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, writing the word cat is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 1 | .1 | .3 | .3 |
| | 1 | 377 | 20.9 | 99.7 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, writing the name Meg is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 2 | .1 | .5 | .5 |
| | 1 | 376 | 20.8 | 99.5 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, writing the word can is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 131 | 7.3 | 34.7 | 34.7 |
| | 1 | 247 | 13.7 | 65.3 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, writing the word walk is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 | 378 | 20.9 | 100.0 | 100.0 |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**For item 5, writing the words to the school, assume the student's response is a word. The score the student should receive is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 39 | 2.2 | 10.3 | 10.3 |
| | 1 | 339 | 18.8 | 89.7 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**In this video the Assessor did something to elicit more writing from the student which was to**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 373 | 20.7 | 98.7 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**On some of the items the Assessor points to where the student should write her responses, this is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 68 | 3.8 | 18.0 | 18.0 |
| | 1 | 310 | 17.2 | 82.0 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1,**

**Write a sentence about what you would name your horse.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 15 | .8 | 4.0 | 4.0 |
| | 1 | 363 | 20.1 | 96.0 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Write a sentence about what you would feed your horse.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 159 | 8.8 | 42.1 | 42.1 |
| | 1 | 219 | 12.1 | 57.9 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Write a sentence about where your horse would live.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 106 | 5.9 | 28.0 | 28.0 |
| | 1 | 272 | 15.1 | 72.0 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Write a sentence on where you would ride your horse.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 93 | 5.1 | 24.6 | 24.6 |
| | 1 | 285 | 15.8 | 75.4 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Write a sentence how you feel riding your horse.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 39 | 2.2 | 10.3 | 10.3 |
| | 1 | 339 | 18.8 | 89.7 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**Scoring an I for Inappropriate for item number 5 for this student would be**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 373 | 20.7 | 98.7 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

**The assessor forgot something in each of these items that is important to do, which was**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 62 | 3.4 | 16.4 | 16.4 |
| | 1 | 316 | 17.5 | 83.6 | 100.0 |
| | Total | 378 | 20.9 | 100.0 | |
| Missing | System | 1428 | 79.1 | | |
| Total | | 1806 | 100.0 | | |

## SCIENCE

The science proficiency assessment evaluates Assessor understanding of the administration and scoring requirements related to the Oregon Extended Assessments in the domain of science. Prerequisite skills tasks and content tasks are represented in the assessment, as are Scaffold and Standard versions of the Oregon Extended Assessments.

**The score the student should receive for item 1, Hello, or Hi.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 10 | .6 | 2.7 | 2.7 |
| | 1 | 365 | 20.2 | 97.3 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Which picture shows cloudy?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 6 | .3 | 1.6 | 1.6 |
| | 1 | 369 | 20.4 | 98.4 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 8, Which is the picture of a leaf from a tree?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 17 | .9 | 4.5 | 4.5 |
| | 1 | 358 | 19.8 | 95.5 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 9, Which is the picture of a microscope?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 9 | .5 | 2.4 | 2.4 |
| | 1 | 366 | 20.3 | 97.6 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 10,**
**Which picture shows something that was made for people to ride?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 9 | .5 | 2.4 | 2.4 |
| | 1 | 366 | 20.3 | 97.6 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1,**
**Which picture shows something you could study: a lamp, candle, or fire?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 7 | .4 | 1.9 | 1.9 |
| | 1 | 368 | 20.4 | 98.1 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2,**
**Which is a tool they could use: a toaster, timer, or television?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 1 | .1 | .3 | .3 |
| | 1 | 374 | 20.7 | 99.7 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Which month on the graph shows the most amount of rainfall: April, May, or June?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 372 | 20.6 | 99.2 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Did his times get slower, faster, or stay the same?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 30 | 1.7 | 8.0 | 8.0 |
| | 1 | 345 | 19.1 | 92.0 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5,**
**Is this because they used a different tool for time, length, or weight?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 7 | .4 | 1.9 | 1.9 |
| | 1 | 368 | 20.4 | 98.1 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1,**
**Which one will get you there the fastest: a bicycle, car, or horse?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 6 | .3 | 1.6 | 1.6 |
| | 1 | 369 | 20.4 | 98.4 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2,**
**Which tool would be best to cut the paper in half: a saw, key, or scissors?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 | 375 | 20.8 | 100.0 | 100.0 |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3,**
**Which cart does it look like will go the fastest: A, B, or C?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 2 | .1 | .5 | .5 |
| | 1 | 373 | 20.7 | 99.5 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4,**
**What data would be best for him to collect on the cart: weight, length, or speed?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 2 | .1 | .5 | .5 |
| | 1 | 373 | 20.7 | 99.5 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Which one of these would cost the least amount of money to make: electric car, lightweight bicycle, or faster airplane?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 1 | .1 | .3 | .3 |
| | 1 | 374 | 20.7 | 99.7 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1,**
**What would you buy to save the most time: dish soap, dishwasher, or dish towel?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 1 | .1 | .3 | .3 |
| | 1 | 374 | 20.7 | 99.7 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, What would be the fastest way to get there, by bicycle, car, or airplane?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 1 | 375 | 20.8 | 100.0 | 100.0 |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Which would the people working with the machine need to protect them from the loud sound?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 5 | .3 | 1.3 | 1.3 |
| | 1 | 370 | 20.5 | 98.7 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, If a stove was made to use less wood, could there be more trees, lakes, or flowers in the future?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 3 | .2 | .8 | .8 |
| | 1 | 372 | 20.6 | 99.2 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Which picture shows an invention that came from the telephone: A, B, or C?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 2 | .1 | .5 | .5 |
| | 1 | 373 | 20.7 | 99.5 | 100.0 |
| | Total | 375 | 20.8 | 100.0 | |
| Missing | System | 1431 | 79.2 | | |
| Total | | 1806 | 100.0 | | |

## REFRESHER

The refresher proficiency assessment evaluates Assessor understanding of the entire assessment system, including assessment options available to Oregon students, the selection of the appropriate assessment option, and administration and scoring expectations for the Prerequisite skills tasks, as well as the Scaffold and Standard versions of the Oregon Extended Assessments, and administration and scoring knowledge and skills related to reading, writing, math, and science. The assessment is used for returning users to continue to demonstration required skills and knowledge.

**For the 2010-11 school year, the Math Extended Assessment will be**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 109 | 6.0 | 10.4 | 10.4 |
| | 1 | 938 | 51.9 | 89.6 | 100.0 |
| | Total | 1047 | 58.0 | 100.0 | |
| Missing | System | 759 | 42.0 | | |
| Total | | 1806 | 100.0 | | |

**For the 2010-11 school year, the high school tests will be for grade/s**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 143 | 7.9 | 13.7 | 13.7 |
| | 1 | 903 | 50.0 | 86.3 | 100.0 |
| | Total | 1046 | 57.9 | 100.0 | |
| Missing | System | 760 | 42.1 | | |
| Total | | 1806 | 100.0 | | |

**For the 2010-11 school year, students taking the Science Extended Assessment in grades 5, 8 and 11 will also have the opportunity to take**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 76 | 4.2 | 7.3 | 7.3 |
| | 1 | 969 | 53.7 | 92.7 | 100.0 |
| | Total | 1045 | 57.9 | 100.0 | |
| Missing | System | 761 | 42.1 | | |
| Total | | 1806 | 100.0 | | |

**Review the Science Field Test Task Selection Table. Mrs. Moore has students to test in grades 5, 8, and 11 this year for Science. She will administer the regular Science Extended Assessment and which t**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 46 | 2.5 | 4.4 | 4.4 |
| | 1 | 997 | 55.2 | 95.6 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The assessment window for 2010-11 Oregon Extended Assessments is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 33 | 1.8 | 3.2 | 3.2 |
| | 1 | 1011 | 56.0 | 96.8 | 100.0 |
| | Total | 1044 | 57.8 | 100.0 | |
| Missing | System | 762 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, How many boxes are shaded?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 20 | 1.1 | 1.9 | 1.9 |
| | 1 | 1023 | 56.6 | 98.1 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, What percent of the six boxes?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 51 | 2.8 | 4.9 | 4.9 |
| | 1 | 992 | 54.9 | 95.1 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, What fraction of the boxes?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 24 | 1.3 | 2.3 | 2.3 |
| | 1 | 1019 | 56.4 | 97.7 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Where would the fraction 1/4 go; and where would the decimal 1.5 go?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 142 | 7.9 | 13.6 | 13.6 |
| | 1 | 901 | 49.9 | 86.4 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, There is one chocolate bar**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 15 | .8 | 1.4 | 1.4 |
| | 1 | 1028 | 56.9 | 98.6 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Which one is a letter?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 57 | 3.2 | 5.5 | 5.5 |
| | 1 | 986 | 54.6 | 94.5 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Where is the letter that makes the mmm sound?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 108 | 6.0 | 10.4 | 10.4 |
| | 1 | 935 | 51.8 | 89.6 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Where is a word?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 52 | 2.9 | 5.0 | 5.0 |
| | 1 | 991 | 54.9 | 95.0 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 6, Where is a word that starts with the letter B?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 94 | 5.2 | 9.0 | 9.0 |
| | 1 | 949 | 52.5 | 91.0 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 8, Where is a story?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 87 | 4.8 | 8.3 | 8.3 |
| | 1 | 956 | 52.9 | 91.7 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, If each of these items were dropped**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 12 | .7 | 1.2 | 1.2 |
| | 1 | 1031 | 57.1 | 98.8 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Which one of these three things?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 7 | .4 | .7 | .7 |
| | 1 | 1036 | 57.4 | 99.3 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Which one of these pictures would require?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 20 | 1.1 | 1.9 | 1.9 |
| | 1 | 1023 | 56.6 | 98.1 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Which one of these requires the most amount of?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 48 | 2.7 | 4.6 | 4.6 |
| | 1 | 995 | 55.1 | 95.4 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Here is the weather forecast for the next week**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 57 | 3.2 | 5.5 | 5.5 |
| | 1 | 986 | 54.6 | 94.5 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 1, Write about where you would go with a car.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 63 | 3.5 | 6.0 | 6.0 |
| | 1 | 980 | 54.3 | 94.0 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 2, Write about the place you go.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 281 | 15.6 | 26.9 | 26.9 |
| | 1 | 762 | 42.2 | 73.1 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 3, Write about whom you meet there.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 303 | 16.8 | 29.1 | 29.1 |
| | 1 | 740 | 41.0 | 70.9 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 4, Write about what will happen there.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 151 | 8.4 | 14.5 | 14.5 |
| | 1 | 892 | 49.4 | 85.5 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

**The score the student should receive for item 5, Write about how you feel about being there.**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 210 | 11.6 | 20.1 | 20.1 |
| | 1 | 833 | 46.1 | 79.9 | 100.0 |
| | Total | 1043 | 57.8 | 100.0 | |
| Missing | System | 763 | 42.2 | | |
| Total | | 1806 | 100.0 | | |

## ADMINISTRATION

The administration proficiency assessment evaluates Assessor understanding of the assessment options available to Oregon students, the selection of the appropriate assessment option, and administration and scoring expectations for the Prerequisite skills tasks, as well as the Scaffold and Standard versions of the Oregon Extended Assessments.

**Oregon Extended Assessments are available for which students?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 9 | .5 | 2.3 | 2.3 |
| | 1 | 380 | 21.0 | 97.7 | 100.0 |
| | Total | 389 | 21.5 | 100.0 | |
| Missing | System | 1417 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**Which list best describes the statewide assessment option(s) currently available for students with IEPs?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 15 | .8 | 3.9 | 3.9 |
| | 1 | 373 | 20.7 | 96.1 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**A fourth grade student taking Extended Assessments would be required to take which of the following assessments?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 24 | 1.3 | 6.2 | 6.2 |
| | 1 | 364 | 20.2 | 93.8 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**Which combination of assessments might an eighth grade student taking Extended Assessments take?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 75 | 4.2 | 19.3 | 19.3 |
| | 1 | 313 | 17.3 | 80.7 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The testing schedule for the Oregon Extended Assessments**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 44 | 2.4 | 11.3 | 11.3 |
| | 1 | 344 | 19.0 | 88.7 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**What is the difference between Scaffold Administration and Standard Administration of the Extended Assessment?**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 79 | 4.4 | 20.4 | 20.4 |
| | 1 | 309 | 17.1 | 79.6 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The Extended Assessments are administered**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 2 | .1 | .5 | .5 |
| | 1 | 386 | 21.4 | 99.5 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**Prerequisite skills**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 75 | 4.2 | 19.3 | 19.3 |
| | 1 | 313 | 17.3 | 80.7 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**When administering the Prerequisite Skills to a student the administrator**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 26 | 1.4 | 6.7 | 6.7 |
| | 1 | 362 | 20.0 | 93.3 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**During the administration of a Prerequisite Skill a score of 3 would be awarded for the student receiving**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 32 | 1.8 | 8.2 | 8.2 |
| | 1 | 356 | 19.7 | 91.8 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The score of "I" is recorded for a student if**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 63 | 3.5 | 16.2 | 16.2 |
| | 1 | 325 | 18.0 | 83.8 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**Content Prompts**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 221 | 12.2 | 57.0 | 57.0 |
| | 1 | 167 | 9.2 | 43.0 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**Scoring in the Content Prompts**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 166 | 9.2 | 42.8 | 42.8 |
| | 1 | 222 | 12.3 | 57.2 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The response choices (often three items) in the Student Materials can be read to students (an allowable accommodation) in each of the subject areas except**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 20 | 1.1 | 5.2 | 5.2 |
| | 1 | 368 | 20.4 | 94.8 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**If a student receives a score of 3 for the Independence for Access**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 115 | 6.4 | 29.6 | 29.6 |
| | 1 | 273 | 15.1 | 70.4 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**For the 2010-11 school year, the Math Extended Assessment will be**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 41 | 2.3 | 10.6 | 10.6 |
| | 1 | 347 | 19.2 | 89.4 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**For the 2010-11 school year, the high school tests will be for grade/s**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 26 | 1.4 | 6.7 | 6.7 |
| | 1 | 362 | 20.0 | 93.3 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**For the 2010-11 school year, students taking the Science Extended Assessment in grades 5, 8 and 11 will also have the opportunity to take**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 41 | 2.3 | 10.6 | 10.6 |
| | 1 | 347 | 19.2 | 89.4 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**Review the Science Field Test Task Selection Table. Mrs. Moore has students to test in grades 5, 8, and 11 this year for Science. She will administer the regular Science Extended Assessment and which one**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 32 | 1.8 | 8.2 | 8.2 |
| | 1 | 356 | 19.7 | 91.8 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

**The assessment window for 2010-11 Oregon Extended Assessments is**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 15 | .8 | 3.9 | 3.9 |
| | 1 | 373 | 20.7 | 96.1 | 100.0 |
| | Total | 388 | 21.5 | 100.0 | |
| Missing | System | 1418 | 78.5 | | |
| Total | | 1806 | 100.0 | | |

## Demographics

The full demographics for students taking the Oregon Extended Assessment are reported in *Appendix A*. Students race/ethnicity was reported in seven categories: (a) Asian/Pacific Islander, (b) American Indian/Alaskan Native, (c) Black, (d) Hispanic, (e) Multiethnic, (f) White, and (g) Decline/Missing. In each grade, the majority of students' ethnic categories were reported as Hispanic or White.

### Reading

***Elementary.*** For grade 3, approximately 69.5% were male, 57.6% were White, and 29.2% were Hispanic. Approximately 66.9% of all students were administered the Standard version of the test, while the remaining 33.1% were administered the Scaffold version of the test. For grade 4, approximately 66.3% were male, 55.1% were White, and 30.9% were Hispanic. Approximately 69.8% of all students were administered the Standard version of the test, while the remaining 30.2% were administered the Scaffold version of the test. For grade 5, 63.7% were male, 58.1% were White, and 26% were Hispanic. Approximately 67.3% of all students were administered the Standard version of the test, while the remaining 32.7% were administered the Scaffold version of the test.

***Middle.*** For grade 6, approximately 67.1% were male, 59% were White, and 25.2% were Hispanic. Approximately 64.2% of all students were administered the Standard version of the test, while the remaining 35.8% were administered the Scaffold version of the test. For grade 7, approximately 65.5% were male, 62% were White, and 25.6% were Hispanic. Approximately 57.9% of all students were administered the Standard version of the test, while the remaining 42.1% were administered the Scaffold version of the test. For grade 8, 65.6% were male, 60.8% were White, and 23.8% were Hispanic. Approximately 56.5% of all students were administered the Standard version of the test, while the remaining 43.5% were administered the Scaffold version of the test.

***High.*** Approximately 65.5% were male, 65.5% were White, and 17.2% were Hispanic. Approximately 43.5% of all students were administered the Standard version of the test, while the remaining 56.5% were administered the Scaffold version of the test.

### Writing

***Grade 4.*** Approximately 68.4% were male, 58% were White, and 26.8% were Hispanic. Approximately 66.2% of all students were administered the Standard version of the test, while the remaining 33.8% were administered the Scaffold version of the test.

***Grade 7.*** Approximately 66.4% were male, 63.5% were White, and 24.3% were Hispanic. Approximately 55.6% of all students were administered the Standard version of the test, while the remaining 44.4% were administered the Scaffold version of the test.

***Grade 11.*** Approximately 65.1% were male, 63.6% were White, and 18.9% were Hispanic. Approximately 42% of all students were administered the Standard version of the test, while the remaining 58% were administered the Scaffold version of the test.

**Math**

***Grade 3.*** Approximately 67% of students taking the mathematics portion of the Oregon Extended Assessment were male, 58% were White, and 29% were Hispanic. Approximately 62% of all students were administered the Standard version of the test, while the remaining 38% were administered the Scaffold version of the test.

***Grade 4.*** Approximately 63% were male, 56% were White, and 29% were Hispanic. Approximately 65% of all students were administered the Standard version of the test, while the remaining 35% were administered the Scaffold version of the test.

***Grade 5.*** Approximately 61% were male, 58% were White, and 26% were Hispanic. Approximately 64% of all students were administered the Standard version of the test, while the remaining 36% were administered the Scaffold version of the test.

***Grade 6.*** Approximately 64% were male, 60% were White, and 24% were Hispanic. Approximately 61% of all students were administered the Standard version of the test, while the remaining 39% were administered the Scaffold version of the test.

***Grade 7.*** Approximately 63% were male, 61% were White, and 26% were Hispanic. Approximately 57% of all students were administered the Standard version of the test, while the remaining 43% were administered the Scaffold version of the test.

***Grade 8.*** Approximately 62% were male, 61% were White, and 23% were Hispanic. Approximately 56% of all students were administered the Standard version of the test, while the remaining 44% were administered the Scaffold version of the test.

***Grade 11.*** Approximately 64% were male, 66% were White, and 17% were Hispanic. Approximately 44% of all students were administered the Standard version of the test, while the remaining 56% were administered the Scaffold version of the test.

**Science**

***Grade 5***. Approximately 61% of students taking the science portion of the Oregon Extended Assessment were male, 61% were White, and 23% were Hispanic. Approximately 57% of all students were administered the Standard version of the test, while the remaining 43% were administered the Scaffold version of the test.

***Grade 8***. Approximately 63% of students taking the science portion of the Oregon Extended Assessment were male, 62% were White, and 23% were Hispanic. Approximately 57% of all students were administered the Standard version of the test, while the remaining 43% were administered the Scaffold version of the test.

***Grade 11.*** Approximately 61% of students taking the science portion of the Oregon Extended Assessment were male, 68% were White, and 17% were Hispanic. Approximately 47% of all students were administered the Standard version of the test, while the remaining 53% were administered the Scaffold version of the test.

## Reliability

Full reliability statistics for the reading portion of the Oregon Extended Assessment are reported in *Appendix B*. These results demonstrate that the total test reliabilities range from moderately high to high (.83 to .97).

### Reading

***Elementary.*** The task reliability for the elementary grade-band (3, 4, 5) was moderate to high, ranging from 0.57 for Task 10 to 0.95 for Task 1. The reliability of the total test was quite high, at 0.94.

Reading: Elementary

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.95 |
| 2 | 0.9 |
| 3 | 0.78 |
| 4 | 0.72 |
| 5 | 0.62 |
| 6 | 0.71 |
| 7 | 0.71 |
| 8 | 0.71 |
| 9 | 0.66 |
| 10 | 0.57 |
| 11 | 0.67 |
| Total Test (Operational Tasks Only) | 0.94 (.92) |

***Middle.*** The task reliability for the middle school grade band (grades 6, 7, and 8) was moderate to high, ranging from 0.59 for Task 7 to 0.96 for Task 1. The reliability of the total test was quite high, at 0.94

Reading: Middle

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.83 |
| 3 | 0.91 |
| 4 | 0.83 |
| 5 | 0.7 |
| 6 | 0.67 |
| 7 | 0.6 |
| 8 | 0.63 |
| 9 | 0.65 |
| 10 | 0.69 |
| 11 | 0.64 |
| Total Test (Operational Tasks Only) | 0.94 (.94) |

***High.*** The task reliability for the high school grade band (grade 11) was moderately high to high, ranging from 0.65 for Task 10 to 0.96 for Task 1. The reliability of the total test was quite high, at 0.96.

Reading: High

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.96 |
| 2 | 0.78 |
| 3 | 0.9 |
| 4 | 0.84 |
| 5 | 0.7 |
| 6 | 0.81 |
| 7 | 0.73 |
| 8 | 0.78 |
| 9 | 0.77 |
| 10 | 0.65 |
| 11 | 0.71 |
| Total Test (Operational Tasks Only) | 0.96 (.95) |

**Writing**

***Grade 4.*** Task reliability was moderate to high, ranging from 0.54 for Task 8 to 0.96 for Task 3. The reliability of the total test was quite high, at 0.93.

Writing: Grade 4

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.95 |
| 2 | 0.93 |
| 3 | 0.96 |
| 4 | 0.7 |
| 5 | 0.78 |
| 6 | 0.64 |
| 7 | 0.7 |
| 8 | 0.54 |
| 9 | 0.85 |
| 10 | 0.61 |
| 11 | 0.79 |
| Total Test (Operational Tasks Only) | 0.93 (.92) |

***Grade 7.*** Task reliability was moderate to high, ranging from 0.68 for Task 7 to 0.97 for Task 2. The reliability of the total test was quite high, at 0.93.

Writing: Grade 7

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.97 |
| 3 | 0.84 |
| 4 | 0.79 |
| 5 | 0.68 |
| 6 | 0.7 |
| 7 | 0.68 |
| 8 | 0.87 |
| 9 | 0.82 |
| 10 | 0.72 |
| 11 | 0.86 |
| Total Test (Operational Tasks Only) | 0.93 (.92) |

***Grade 11.*** Task reliability was moderately high to high, ranging from 0.78 for Task 4 to 0.96 for Task 1. The reliability of the total test was quite high, at 0.97.

Writing: Grade 11

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.97 |
| 3 | 0.89 |
| 4 | 0.78 |
| 5 | 0.84 |
| 6 | 0.81 |
| 7 | 0.86 |
| 8 | 0.92 |
| 9 | 0.88 |
| 10 | 0.81 |
| 11 | 0.87 |
| Total Test (Operational Tasks Only) | 0.97 (.96) |

**Math**

***Grade 3.*** Task reliability was moderate to high, ranging from 0.51 for Task 6 to 0.95 for Task 1. The reliability of the total test was quite high, at 0.92.

Math: Grade 3

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.95 |
| 2 | 0.71 |
| 3 | 0.62 |
| 4 | 0.62 |
| 5 | 0.56 |
| 6 | 0.51 |
| 7 | 0.66 |
| 8 | 0.59 |
| 9 | 0.55 |
| Total Test (Operational Tasks Only) | 0.92 (.83) |

***Grade 4.*** Task reliability was again moderate to high, ranging from 0.51 for Task 6 to 0.95 for Task 1. The reliability of the total test was quite high, at 0.92.

Math Grade 4

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.95 |
| 2 | 0.71 |
| 3 | 0.62 |
| 4 | 0.62 |
| 5 | 0.56 |
| 6 | 0.51 |
| 7 | 0.66 |
| 8 | 0.59 |
| 9 | 0.55 |
| Total Test (Operational Tasks Only) | 0.92 (.88) |

***Grade 5.*** Task reliability was relatively low to high, ranging from 0.36 for Task 7 to 0.96 for Task 1. The reliability of the total test was quite high, at 0.91.

Math: Grade 5

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.96 |
| 2 | 0.56 |
| 3 | 0.65 |
| 4 | 0.56 |
| 5 | 0.56 |
| 6 | 0.53 |
| 7 | 0.36 |
| 8 | 0.48 |
| 9 | 0.51 |
| Total Test (Operational Tasks Only) | 0.91 (.85) |

***Grade 6.*** Task reliability was low to high, ranging from 0.30 for Task 8 to 0.97 for Task 1. The reliability of the total test was moderately high, at 0.86.

Math: Grade 6

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.55 |
| 3 | 0.66 |
| 4 | 0.41 |
| 5 | 0.45 |
| 6 | 0.36 |
| 7 | 0.39 |
| 8 | 0.3 |
| 9 | 0.44 |
| Total Test (Operational Tasks Only) | 0.86 (.79) |

***Grade 7.*** Task reliability was low to high, ranging from 0.24 for Task 7 to 0.97 for Task 1. The reliability of the total test was moderately high, at 0.87.

Math: Grade 7

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.65 |
| 3 | 0.62 |
| 4 | 0.62 |
| 5 | 0.48 |
| 6 | 0.5 |
| 7 | 0.24 |
| 8 | 0.36 |
| 9 | 0.43 |
| Total Test (Operational Tasks Only) | 0.87 (.83) |

***Grade 8.*** Task reliability was low to high, ranging from 0.32 for Task 9 to 0.97 for Task 1. The reliability of the total test was quite high, at 0.88.

Math: Grade 8

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.56 |
| 3 | 0.52 |
| 4 | 0.54 |
| 5 | 0.59 |
| 6 | 0.6 |
| 7 | 0.46 |
| 8 | 0.43 |
| 9 | 0.32 |
| Total Test (Operational Tasks Only) | 0.88 (.85) |

***Grade 11.*** Task reliability was relatively low to high, ranging from 0.29 for Task 5 to 0.97 for Task 1. The reliability of the total test was quite high, at 0.89.

Math Grade 11

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.63 |
| 3 | 0.57 |
| 4 | 0.6 |
| 5 | 0.29 |
| 6 | 0.64 |
| 7 | 0.4 |
| 8 | 0.64 |
| 9 | 0.47 |
| Total Test (Operational Tasks Only) | 0.89 (.87) |

**Science**

***Grade 5***. Task reliability for the operational items was moderate to high, ranging from 0.51 for Task 8 to 0.97 for Task 1. The reliability of the total operational test was quite high, at 0.93. The reliability of the total field test items was also quite high, at 0.98.

Science Operational: Grade 5

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.97 |
| 2 | 0.76 |
| 3 | 0.83 |
| 4 | 0.69 |
| 5 | 0.7 |
| 6 | 0.75 |
| 7 | 0.57 |
| 8 | 0.51 |
| 9 | 0.78 |
| Total Test | 0.93 |

Science Field Test: Grade 5

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.61 |
| 2 | 0.64 |
| 3 | 0.6 |
| 4 | 0.76 |
| 5 | 0.74 |
| 6 | 0.56 |
| 7 | 0.76 |
| 8 | 0.63 |
| Total Field Test | 0.98 |

***Grade 8.*** Task reliability for the operational items was moderately low to high, ranging from 0.39 for Task 5 to 0.98 for Task 1. The reliability of the total operational test was quite high, at 0.90. The reliability of the total field test items was also quite high, at 0.93.

Science Operational: Grade 8

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.98 |
| 2 | 0.62 |
| 3 | 0.74 |
| 4 | 0.64 |
| 5 | 0.39 |
| 6 | 0.54 |
| 7 | 0.58 |
| 8 | 0.48 |
| 9 | 0.68 |
| Total Test | 0.90 |

Science Field Test: Grade 8

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.61 |
| 2 | 0.64 |
| 3 | 0.59 |
| 4 | 0.74 |
| 5 | 0.4 |
| 6 | 0.34 |
| 7 | 0.69 |
| 8 | 0.51 |
| Total Field Test | 0.93 |

***Grade 11.*** Task reliability for the operational items was moderately low to high, ranging from 0.39 for Task 7 to 0.98 for Task 1. The reliability of the total operational test was quite high, at 0.88. The reliability of the total field test items was also quite high, at 0.98.

Science Operational: Grade 11

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.98 |
| 2 | 0.66 |
| 3 | 0.74 |
| 4 | 0.65 |
| 5 | 0.54 |
| 6 | 0.39 |
| 7 | 0.39 |
| 8 | 0.65 |
| 9 | 0.54 |
| Total Test | 0.89 |

Science Field Test: Grade 11

| Task | Cronbach's Alpha |
|---|---|
| 1 | 0.69 |
| 2 | 0.64 |
| 3 | 0.59 |
| 4 | 0.63 |
| 5 | 0.7 |
| 6 | 0.79 |
| 7 | 0.69 |
| 8 | 0.76 |
| Total Field Test | 0.98 |

## Descriptive Statistics

The Oregon Extended Assessments are part of a large-scale assessment system that is developed, administered, scored, and reported in concert with the professional expectations established by the *Standards*[8] (AERA et al., 1999) and best professional practices. Items are developed in an iterative manner that includes evaluation by Oregon teachers and education professionals for bias, accessibility, and alignment to the appropriate Oregon standards.

The assessments evaluate a level of student performance that has been reduced in terms of depth, breadth, and complexity in comparison to Oregon's content standards. These assessments reflect an appropriate range of performance demands (easy to difficult) to assess students with significant cognitive disabilities who exhibit a wide variety of achievement levels.

Full descriptive statistics for the reading items of the Oregon Extended Assessment are reported in *Appendix C*. All Tasks 1 were scored on a 4-point scale. All subsequent Tasks were scored on a 2-point scale. In general, the test has an appropriate range of item difficulties represented, from easy to difficult. The easiest items are located in Task 1, the prerequisite skills items. Item difficulties range from $p$=.15 (the most difficult item) to $p$=.96 (the easiest item). Item difficulties are deemed appropriate across all subject areas.

### Reading: Elementary

***Task 1***. Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.84 – 0.96. Generally, the less difficult items had a lower standard deviation. For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2***. Items were relatively easy overall. Item 2 was the most difficult item, $p$ = 0.64, while item 3 was the easiest, $p$ = 0.85. Item 3 had the lowest standard deviation (0.68) while Item 1 had the highest (0.76). Students averaged a total score of 7.39 with a standard deviation of 3.30.

***Task 3 (field test).*** Items overall were relatively easy. Item 5 was the most difficult item, $p$ = 0.65, while item 3 was the easiest, $p$ = 0.83. Items 2 and 3 had the lowest standard deviation (0.74) while Item 5 had the highest (0.80). Students averaged a total score of 7.46 with a standard deviation of 2.94.

***Task 4 (field test).*** Item 2 was the most difficult item, $p$ = 0.72, while item 1 was the easiest, $p$ = 0.84. Item 1 had the lowest standard deviation (0.63) while Item 4 had the highest (0.71). Students averaged a total score of 7.32 with a standard deviation of 3.06.

***Task 5.*** Item 1 was the most difficult item, $p$ = 0.68, while item 2 was the easiest, $p$ = 0.83. Item 5 had the lowest standard deviation (0.70) while Item 1 had the highest (0.93). Students averaged a total score of 7.11 with a standard deviation of 3.24.

***Task 6.*** Item 5 was the most difficult item, $p$ = 0.74, while item 3 was the easiest, $p$ = 0.92. Item 2 had the lowest standard deviation (0.50) while Items 5 had the highest (0.71). Students averaged a total score of 7.47 with a standard deviation of 3.17.

***Task 7.*** Item 5 was the most difficult item, $p$ = 0.68, while item 1 was the easiest, $p$ = 0.90. Item 1 had the lowest standard deviation (0.52) while Item 5 had the highest (0.78). Students averaged a total score of 7.32 with a standard deviation of 3.31.

***Task 8.*** Item 5 was the most difficult item, $p$ = 0.69, while item 4 was the easiest, $p$ = 0.88. Item 4 had the lowest standard deviation (0.61) while Item 5 had the highest (0.92). Students averaged a total score of 6.68 with a standard deviation of 3.38.

***Task 9.*** Item 5 was the most difficult item, $p$ = 0.69, while item 4 was the easiest, $p$ = 0.88. Item 4 had the lowest standard deviation (0.61) while Item 5 had the highest (0.92). Students averaged a total score of 6.81 with a standard deviation of 3.42.

---

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

***Task 10.*** Item 5 was the most difficult item, $p = 0.59$, while item 1 was the easiest, $p = 0.77$. Item 1 had the lowest standard deviation (0.68) while Item 5 had the highest (0.87). Students averaged a total score of 6.00 with a standard deviation of 3.37.

***Task 11.*** Item 5 was the most difficult item, $p = 0.65$, while item 1 was the easiest, $p = 0.85$. Item 5 had the lowest standard deviation (0.65) while Item 4 had the highest (0.69). Students averaged a total score of 6.34 with a standard deviation of 3.27.

***Total test.*** The average total test score was 69.90 with a standard deviation of 28.43. Item difficulties range from $p$=.59 (the most difficult item) to $p$=.96 (the easiest item).

**Reading: Middle**

***Task 1***. Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.84 – 0.95. Item 1 had the lowest standard deviation (0.61) and item 10 had the highest (0.94). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** Items in task 2 were more difficult in general, compared to Task 1. Item 4 was the most difficult item, $p = 0.70$, while item 3 was the easiest, $p = 0.80$. Item 3 had the lowest standard deviation (0.74) while Item 2 had the highest (0.85). Students averaged a total score of 7.15 with a standard deviation of 3.27.

***Task 3.*** Item 4 was the most difficult item, $p = 0.66$, while items 1 and 2 were the easiest, $p = 0.70$. Item 1 had the lowest standard deviation (0.83) while Item 4 had the highest (0.91). Students averaged a total score of 6.58 with a standard deviation of 3.90.

***Task 4.*** Item 4 was the most difficult item, $p = 0.55$, while item 5 was the easiest, $p = 0.80$. Item 5 had the lowest standard deviation (0.69) while Item 4 had the highest (0.95). Students averaged a total score of 6.16 with a standard deviation of 3.75.

***Task 5.*** Item 2 was the most difficult item, $p = 0.73$, while items 3 and 5 were the easiest, $p = 0.83$. Item 4 had the lowest standard deviation (0.66) while Item 5 had the highest (0.67). Students averaged a total score of 6.58 with a standard deviation of 3.56.

***Task 6 (field test).*** Item 3 was the most difficult item, $p = 0.77$, while item 2 was the easiest, $p = 0.89$. Item 2 had the lowest standard deviation (0.56) while Item 3 had the highest (0.74). Students averaged a total score of 6.96 with a standard deviation of 3.71.

***Task 7.*** Item 2 was the most difficult item, $p = 0.66$, while item 1 was the easiest, $p = 0.87$. Item 1 had the lowest standard deviation (0.56) while Item 2 had the highest (0.88). Students averaged a total score of 6.29 with a standard deviation of 3.54.

***Task 8.*** Item 5 was the most difficult item, $p = 0.68$, while item 3 was the easiest, $p = 0.91$. Item 3 had the lowest standard deviation (0.54) while Item 5 had the highest (0.68). Students averaged a total score of 6.46 with a standard deviation of 3.56.

***Task 9.*** Item 5 was the most difficult item, $p = 0.74$, while item 1 was the easiest, $p = 0.90$. Item 1 had the lowest standard deviation (0.51) while Item 5 had the highest (0.76). Students averaged a total score of 6.74 with a standard deviation of 3.74.

***Task 10.*** Item 3 was the most difficult item, $p = 0.71$, while item 2 was the easiest, $p = 0.85$. Item 2 had the lowest standard deviation (0.58) while Item 4 had the highest (0.71). Students averaged a total score of 6.40 with a standard deviation of 3.63.

***Task 11.*** Item 3 was the most difficult item, $p = 0.57$, while item 2 was the easiest, $p = 0.85$. Item 2 had the lowest standard deviation (0.52) while Item 4 had the highest (0.74). Students averaged a total score of 5.88 with a standard deviation of 3.41.

***Total test.*** The average total test score was 65.22 with a standard deviation of 33.26. Item difficulties range from $p$=.55 (the most difficult item) to $p$=.95 (the easiest item).

**Reading: High**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.80 – 0.96. Item 1 had the lowest standard deviation (0.56) and item 10 had the highest (1.03). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** Items in task 2 were more difficult in general, compared to Task 1. Item 2 was the most difficult item, $p = 0.41$, while item 4 was the easiest, $p = 0.70$. Item 4 had the lowest standard deviation (0.81) while Item 5 had the highest (0.87). Students averaged a total score of 5.99 with a standard deviation of 3.25.

***Task 3.*** Item 5 was the most difficult item, $p = 0.52$, while items 1 was the easiest, $p = 0.65$. Item 5 had the lowest standard deviation (0.87) while Item 4 had the highest (0.94). Students averaged a total score of 5.66 with a standard deviation of 3.98.

***Task 4 (field test).*** Item 4 was the most difficult item, $p = 0.62$, while item 3 was the easiest, $p = 0.77$. Item 3 had the lowest standard deviation (0.70) while Item 4 had the highest (0.86). Students averaged a total score of 5.61 with a standard deviation of 3.79.

***Task 5 (field test).*** Item 5 was the most difficult item, $p = 0.53$, while item 2 was the easiest, $p = 0.80$. Item 2 had the lowest standard deviation (0.71) while Item 1 had the highest (0.80). Students averaged a total score of 5.45 with a standard deviation of 3.71.

***Task 6.*** Item 3 was the most difficult item, $p = 0.76$, while item 5 was the easiest, $p = 0.86$. Item 5 had the lowest standard deviation (0.61) while Item 2 had the highest (0.73). Students averaged a total score of 6.12 with a standard deviation of 4.10.

***Task 7.*** Item 5 was the most difficult item, $p = 0.49$, while item 1 was the easiest, $p = 0.84$. Item 4 had the lowest standard deviation (0.64) while Item 5 had the highest (0.94). Students averaged a total score of 5.28 with a standard deviation of 3.78.

***Task 8.*** Item 5 was the most difficult item, $p = 0.70$, while item 3 was the easiest, $p = 0.83$. Item 2 had the lowest standard deviation (0.62) while Item 5 had the highest (0.80). Students averaged a total score of 5.67 with a standard deviation of 4.11.

***Task 9.*** Item 5 was the most difficult item, $p = 0.70$, while item 1 was the easiest, $p = 0.87$. Item 1 had the lowest standard deviation (0.58) while Item 5 had the highest (0.79). Students averaged a total score of 5.67 with a standard deviation of 3.96.

***Task 10.*** Item 2 was the most difficult item, $p = 0.62$, while item 4 was the easiest, $p = 0.78$. Items 1 and 5 had the lowest standard deviation (0.67) while Item 3 had the highest (0.70). Students averaged a total score of 5.12 with a standard deviation of 3.63.

***Task 11.*** Items 1 and 2 were the most difficult items, $p = 0.59$, while item 4 was the easiest, $p = 0.68$. Item 1 had the lowest standard deviation (0.60) while Item 5 had the highest (0.69). Students averaged a total score of 4.64 with a standard deviation of 3.27.

***Total test.*** The average total test score was 55.21 with a standard deviation of 34.81. Item difficulties range from $p$=.41 (the most difficult item) to $p$=.96 (the easiest item).

**Writing: Grade 4**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.89 – 0.96. Item 1 had the lowest standard deviation (0.55) while Item 8 had the highest (0.97). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** Items in task 2 were relatively easy overall. Item 4 was the most difficult item, $p = 0.65$, while item 1 was the easiest, $p = 0.85$. Item 1 had the lowest standard deviation (0.66) while Item 4 had the highest (0.79). Students averaged a total score of 7.41 with a standard deviation of 3.40.

***Task 3.*** Item 4 was the most difficult item, $p = 0.66$, while item 1 was the easiest, $p = 0.82$. Item 1 had the lowest standard deviation (0.70) while Item 5 had the highest (0.75). Students averaged a total score of 6.88 with a standard deviation of 3.61.

***Task 4.*** Item 5 was the most difficult item, $p = 0.56$, while item 1 was the easiest, $p = 0.81$. Item 1 had the lowest standard deviation (0.62) while Item 4 had the highest (0.72). Students averaged a total score of 5.84 with a standard deviation of 3.26.

***Task 5.*** Item 5 was the most difficult item, $p = 0.68$, while item 3 was the easiest, $p = 0.83$. Item 3 had the lowest standard deviation (0.63) while Item 5 had the highest (0.69). Students averaged a total score of 6.55 with a standard deviation of 3.53.

***Task 6.*** Item 1 was the most difficult item, $p = 0.50$, while item 4 was the easiest, $p = 0.81$. Item 4 had the lowest standard deviation (0.64) while Items 5 had the highest (0.88). Students averaged a total score of 5.18 with a standard deviation of 3.36.

***Task 7.*** Item 3 was the most difficult item, $p = 0.48$, while item 1 was the easiest, $p = 0.62$. Item 1 had the lowest standard deviation (0.93) while Item 5 had the highest (0.97). Students averaged a total score of 4.23 with a standard deviation of 3.61.

***Task 8 (field test).*** Item 4 was the most difficult item, $p = 0.56$, while items 1 and 5 were the easiest, $p = 0.64$. Item 3 had the lowest standard deviation (0.55) while Item 1 had the highest (0.91). Students averaged a total score of 4.59 with a standard deviation of 3.08.

***Task 9.*** Item 5 was the most difficult item, $p = 0.37$, while item 1 was the easiest, $p = 0.57$. Item 1 had the lowest standard deviation (0.75) while Item 5 had the highest (0.78). Students averaged a total score of 3.16 with a standard deviation of 3.27.

***Task 10 (field test).*** Items in task 10 were quite difficult compared to other tasks. Item 1 was the most difficult item, $p = 0.25$, while item 3 was the easiest, $p = 0.60$. Item 3 had the lowest standard deviation (0.64) while Item 2 had the highest (0.73). Students averaged a total score of 2.85 with a standard deviation of 2.56.

***Task 11.*** Item 1 was the most difficult item, $p = 0.67$, while item 5 was the easiest, $p = 0.94$. Item 5 had the lowest standard deviation (0.42) while Item 3 had the highest (0.82). Students averaged a total score of 5.67 with a standard deviation of 3.92.

***Total test.*** The average total test score was 44.31 with a standard deviation of 33.49. Item difficulties range from $p=.25$ (the most difficult item, a field test item) to $p=.96$ (the easiest item).

**Writing: Grade 7**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties (*p* value) ranging from 0.86 – 0.96. Item 1 had the lowest standard deviation (0.60) and item 8 had the highest (1.10). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** Items in task 2 were more difficult in general, compared to Task 1. Item 5 was the most difficult item, $p = 0.69$, while items 1 and 4 were the easiest, $p = 0.75$. Items 1 and 2 had the lowest standard deviation (0.75) while Items 3 and 4 had the highest (0.77). Students averaged a total score of 6.75 with a standard deviation of 3.98.

***Task 3.*** Item 5 was the most difficult item, $p = 0.59$, while items 1 was the easiest, $p = 0.77$. Items 1 and 2 had the lowest standard deviation (0.77) while Item 5 had the highest (0.86). Students averaged a total score of 6.17 with a standard deviation of 3.48.

***Task 4.*** Item 4 was the most difficult item, $p = 0.63$, while item 2 was the easiest, $p = 0.85$. Item 2 had the lowest standard deviation (0.64) while Item 5 had the highest (0.67). Students averaged a total score of 6.06 with a standard deviation of 3.44.

***Task 5.*** Item 3 was the most difficult item, $p = 0.67$, while item 5 was the easiest, $p = 0.87$. Items 3 and 5 had the lowest standard deviation (0.61) while Item 2 had the highest (0.76). Students averaged a total score of 6.02 with a standard deviation of 3.70.

***Task 6 (field test).*** Item 2 was the most difficult item, $p = 0.75$, while item 4 was the easiest, $p = 0.87$. Item 5 had the lowest standard deviation (0.54) while Item 2 had the highest (0.77). Students averaged a total score of 6.37 with a standard deviation of 3.98.

***Task 7.*** Item 5 was the most difficult item, $p = 0.52$, while item 2 was the easiest, $p = 0.76$. Item 1 had the lowest standard deviation (0.58) while Item 5 had the highest (0.88). Students averaged a total score of 5.07 with a standard deviation of 3.52.

***Task 8.*** Item 5 was the most difficult item, $p = 0.55$, while item 1 was the easiest, $p = 0.66$. Item 1 had the lowest standard deviation (0.71) while Item 5 had the highest (0.79). Students averaged a total score of 3.95 with a standard deviation of 3.72.

***Task 9.*** Item 4 was the most difficult item, $p = 0.59$, while item 3 was the easiest, $p = 0.94$. Item 3 had the lowest standard deviation (0.42) while Items 2, 4, and 5 had the highest (0.76). Students averaged a total score of 3.90 with a standard deviation of 3.74.

***Task 10.*** Item 3 was the most difficult item, $p = 0.71$, while item 2 was the easiest, $p = 0.85$. Item 2 had the lowest standard deviation (0.58) while Item 4 had the highest (0.71). Students averaged a total score of 4.61 with a standard deviation of 3.82.

***Task 11.*** Item 1 was the most difficult item, $p = 0.68$, while item 5 was the easiest, $p = 0.90$. Item 5 had the lowest standard deviation (0.56) while Item 3 had the highest (0.79). Students averaged a total score of 4.89 with a standard deviation of 4.28.

***Total test.*** The average total test score was 53.79 with a standard deviation of 33.32. Item difficulties range from $p=.52$ (the most difficult item) to $p=.96$ (the easiest item).

**Writing: Grade 11**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties (*p* value) ranging from 0.86 – 0.96. Item 1 had the lowest standard deviation (0.61) and item 8 had the highest (1.08). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** Items in task 2 were more difficult in general, compared to Task 1. Item 3 was the most difficult item, $p = 0.63$, while item 1 was the easiest, $p = 0.74$. Item 1 had the lowest standard deviation (0.76) while Item 4 had the highest (0.78). Students averaged a total score of 6.24 with a standard deviation of 4.01.

***Task 3.*** Item 2 was the most difficult item, $p = 0.48$, while items 3 was the easiest, $p = 0.56$. Item 2 had the lowest standard deviation (0.78) while Items 1 and 5 had the highest (0.89). Students averaged a total score of 4.89 with a standard deviation of 3.67.

***Task 4.*** Item 4 was the most difficult item, $p = 0.39$, while item 1 was the easiest, $p = 0.64$. Item 5 had the lowest standard deviation (0.74) while Item 3 had the highest (0.82). Students averaged a total score of 4.26 with a standard deviation of 3.31.

***Task 5.*** Item 5 was the most difficult item, $p = 0.63$, while item 4 was the easiest, $p = 0.83$. Item 1 had the lowest standard deviation (0.68) while Item 2 had the highest (0.82). Students averaged a total score of 5.44 with a standard deviation of 4.04.

***Task 6 (field test).*** Item 5 was the most difficult item, $p = 0.55$, while item 2 was the easiest, $p = 0.73$. Item 1 had the lowest standard deviation (0.73) while Item 3 had the highest (0.80). Students averaged a total score of 4.68 with a standard deviation of 3.77.

***Task 7 (field test).*** Item 4 was the most difficult item, $p = 0.56$, while item 5 was the easiest, $p = 0.72$. Item 1 had the lowest standard deviation (0.68) while Item 4 had the highest (0.85). Students averaged a total score of 4.58 with a standard deviation of 3.91.

***Task 8.*** Item 5 was the most difficult item, $p = 0.43$, while item 1 was the easiest, $p = 0.57$. Item 1 had the lowest standard deviation (0.80) while Item 5 had the highest (0.83). Students averaged a total score of 2.89 with a standard deviation of 3.63.

***Task 9.*** Item 5 was the most difficult item, $p = 0.41$, while item 3 was the easiest, $p = 0.76$. Item 2 had the lowest standard deviation (0.75) while Item 5 had the highest (0.82). Students averaged a total score of 3.21 with a standard deviation of 3.71.

***Task 10.*** Item 5 was the most difficult item, $p = 0.57$, while item 3 was the easiest, $p = 0.90$. Item 3 had the lowest standard deviation (0.53) while Item 2 had the highest (0.81). Students averaged a total score of 4.04 with a standard deviation of 3.94.

***Task 11***. Item 4 was the most difficult items, $p = 0.55$, while item 5 was the easiest, $p = 0.86$. Item 5 had the lowest standard deviation (0.64) while Item 3 had the highest (0.81). Students averaged a total score of 4.08 with a standard deviation of 4.13.

***Total test.*** The average total test score was 44.31 with a standard deviation of 33.49. Item difficulties range from $p$=.39 (the most difficult item) to $p$=.96 (the easiest item).

**Math: Grade 3**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.85 – 0.94. Generally, the more difficult items had a higher standard deviation than the less difficult items. For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The item range was much larger for task 2 and the items overall were more difficult. Item 5 was the most difficult item, $p = 0.44$, while item 6 (field test) was the easiest, $p = 0.75$. Item 5 had the lowest standard deviation (0.74) while item 2 had the highest (0.98). Students averaged a total task score of 7.78 with a standard deviation of 3.64.

***Task 3.*** Item 6 (field test) was the most difficult item, $p = 0.28$, while item 4 was the easiest, $p = 0.84$. Item 1 had the lowest standard deviation (0.78) while item 3 had the highest (0.98). Students averaged a total task score of 5.06 with a standard deviation of 3.04.

***Task 4.*** Item 5 was the most difficult item, $p = 0.40$, while items 4 and 6 (field test) were the easiest, $p = 0.71$. Item 5 had the lowest standard deviation (0.74) while items 1 and 3 had the highest (0.98). Students averaged a total task score of 4.54 with a standard deviation of 2.66.

***Task 5.*** Item 3 was the most difficult item, $p = 0.39$, while item 5 was the easiest, $p = 0.73$. Item 5 had the lowest standard deviation (0.81) while item 6 (field test) had the highest (1.00). Students averaged a total task score of 5.96 with a standard deviation of 3.32.

***Task 6.*** Item 1 was the most difficult item, $p = 0.41$, while item 6 (field test) was the easiest, $p = 0.83$. Item 6 (field test) had the lowest standard deviation (0.75) while item 2 had the highest (1.00). Students averaged a total task score of 7.26 with a standard deviation of 2.97.

***Task 7.*** Item 1 was the most difficult item, $p = 0.15$, while item 6 (field test) was the easiest, $p = 0.87$. Item 6 (field test) had the lowest standard deviation (0.67) while item 3 had the highest (0.99). Students averaged a total task score of 7.04 with a standard deviation of 2.69.

***Task 8.*** Item 1 was the most difficult item, $p = 0.26$, while item 6 (field test) was the easiest, $p = 0.91$. Item 6 (field test) had the lowest standard deviation (0.59) while item 2 had the highest (0.95). Students averaged a total task score of 7.63 with a standard deviation of 2.79.

***Task 9.*** Item 5 was the most difficult item, $p = 0.41$, while item 3 was the easiest, $p = 0.86$. Item 3 had the lowest standard deviation (0.70) while item 6 (field test) had the highest (0.99). Students averaged a total task score of 7.87 with a standard deviation of 2.76.

***Total test.*** The average total test score was 46.68 with a standard deviation of 23.14. Item difficulties range from $p$=.15 (the most difficult item) to $p$=.94 (the easiest item).

**Math: Grade 4**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.88 – 0.96. Item 1 had the lowest standard deviation (0.57), while item 5 had the highest standard deviation (.91). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The item range was much larger for task 2 and the items overall were more difficult. Item 6 (field test) was the most difficult item, $p = 0.29$, while item 4 was the easiest, $p = 0.77$. Item 2 had the lowest standard deviation (0.78) while item 3 had the highest (0.99). Students averaged a total task score of 6.87 with a standard deviation of 3.41.

***Task 3.*** Item 6 (field test) was the most difficult item, $p = 0.24$, while item 1 was the easiest, $p = 0.67$. Item 3 had the lowest standard deviation (0.71) while item 4 had the highest (1.00). Students averaged a total task score of 5.48 with a standard deviation of 3.24.

***Task 4.*** Item 6 (field test) was the most difficult item, $p = 0.15$, while item 3 was the easiest, $p = 0.71$. Item 5 had the lowest standard deviation (0.72) while item 1 had the highest (0.98). Students averaged a total task score of 7.00 with a standard deviation of 3.10.

***Task 5.*** Item 6 (field test) was the most difficult item, $p = 0.42$, while item 4 was the easiest, $p = 0.74$. Item 3 had the lowest standard deviation (0.73) while item 5 had the highest (0.99). Students averaged a total task score of 7.02 with a standard deviation of 3.11.

***Task 6.*** Item 3 was the most difficult item, $p = 0.45$, while item 6 (field test) was the easiest, $p = 0.83$. Item 6 (field test) had the lowest standard deviation (0.73) while items 1 and 4 had the highest (1.00). Students averaged a total task score of 7.00 with a standard deviation of 3.10.

***Task 7.*** Item 5 was the most difficult item, $p = 0.37$, while item 2 was the easiest, $p = 0.85$. Item 2 had the lowest standard deviation (0.72) while item 4 had the highest (0.99). Students averaged a total task score of 8.07 with a standard deviation of 3.02.

***Task 8.*** Item 1 was the most difficult item, $p = 0.49$, while item 6 (field test) was the easiest, $p = 0.82$. Item 1 had the lowest standard deviation (0.71) while item 3 had the highest (1.00). Students averaged a total task score of 7.44 with a standard deviation of 3.09.

***Task 9.*** Item 4 was the most difficult item, $p = 0.44$, while item 1 was the easiest, $p = 0.79$. Item 3 had the lowest standard deviation (0.75) while item 6 (field test) had the highest (1.00). Students averaged a total task score of 6.63 with a standard deviation of 3.06.

***Total test.*** The average total test score was 48.54 with a standard deviation of 25.54. Item difficulties range from $p$=.15 (the most difficult item) to $p$=.96 (the easiest item).

**Math: Grade 5**

***Task 1***. Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.88 – 0.96. Item 1 had the lowest standard deviation (0.56), while item 5 had the highest standard deviation (.89). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The item range was much larger in task 2 and the items overall were more difficult. Item 2 was the most difficult item, $p = 0.19$, while item 6 (field test) was the easiest, $p = 0.68$. Item 2 had the lowest standard deviation (0.78) while item 1 had the highest (0.99). Students averaged a total task score of 5.48 with a standard deviation of 3.09.

***Task 3.*** Item 3 was the most difficult item, $p = 0.33$, while item 6 (field test) was the easiest, $p = 0.71$. Item 6 (field test) had the lowest standard deviation (0.90) while item 4 had the highest (0.99). Students averaged a total task score of 6.00 with a standard deviation of 3.44.

***Task 4.*** Item 3 was the most difficult item, $p = 0.33$, while item 2 was the easiest, $p = 0.76$. Item 5 had the lowest standard deviation (0.71) while item 1 had the highest (0.97). Students averaged a total task score of 6.47 with a standard deviation of 2.96.

***Task 5.*** Item 6 (field test) was the most difficult item, $p = 0.38$, while item 2 was the easiest, $p = 0.80$. Item 2 had the lowest standard deviation (0.81) while item 3 had the highest (1.00). Students averaged a total task score of 6.44 with a standard deviation of 3.26.

***Task 6.*** Item 5 was the most difficult item, $p = 0.36$, while item 6 (field test) was the easiest, $p = 0.77$. Item 6 (field test) had the lowest standard deviation (0.85) while item 2 had the highest (1.00). Students averaged a total task score of 6.85 with a standard deviation of 3.03.

***Task 7.*** Item 5 was the most difficult item, $p = 0.37$, while item 6 (field test) was the easiest, $p = 0.83$. Item 3 had the lowest standard deviation (0.75) while item 2 had the highest (1.00). Students averaged a total task score of 6.09 with a standard deviation of 2.72.

***Task 8.*** Item 3 was the most difficult item, $p = 0.33$, while item 6 (field test) was the easiest, $p = 0.87$. Item 6 (field test) had the lowest standard deviation (0.67) while item 2 had the highest (0.98). Students averaged a total task score of 6.89 with a standard deviation of 2.75.

***Task 9.*** Item 2 was the most difficult item, $p = 0.41$, while items 1 and 4 were the easiest, $p = 0.80$. Item 6 (field test) had the lowest standard deviation (0.63) while item 3 had the highest (1.00). Students averaged a total task score of 8.21 with a standard deviation of 2.76.

***Total test.*** The average total test score was 45.74 with a standard deviation of 24.12. Item difficulties range from $p$=.19 (the most difficult item) to $p$=.96 (the easiest item).

**Math: Grade 6**

***Task 1***. Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.87 – 0.96. Item 1 had the lowest standard deviation (0.58), while item 5 had the highest standard deviation (.94). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The item range was much larger in task 2 and the items overall were more difficult. Item 2 was the most difficult item, $p = 0.33$, while item 4 was the easiest, $p = 0.57$. Item 5 had the lowest standard deviation (0.81) while item 1 had the highest (1.00). Students averaged a total task score of 5.37 with a standard deviation of 3.19.

***Task 3.*** Item 1 was the most difficult item, $p = 0.37$, while item 5 was the easiest, $p = 0.62$. Item 1 had the lowest standard deviation (0.97) while items 2 and 4 had the highest (0.99). Students averaged a total task score of 6.24 with a standard deviation of 3.60.

***Task 4***. Item 3 was the most difficult item, $p = 0.22$, while item 1 was the easiest, $p = 0.52$. Item 3 had the lowest standard deviation (0.82) while items 1 and 2 had the highest (1.00). Students averaged a total task score of 4.08 with a standard deviation of 2.78.

***Task 5***. Item 6 (field test) was the most difficult item, $p = 0.32$, while item 5 was the easiest, $p = 0.60$. Item 6 (field test) had the lowest standard deviation (0.93) while item 2 had the highest (1.00). Students averaged a total task score of 5.13 with a standard deviation of 3.03.

***Task 6.*** Item 2 was the most difficult item, $p = 0.29$, while item 4 was the easiest, $p = 0.76$. Item 1 had the lowest standard deviation (0.84) while item 3 had the highest (0.99). Students averaged a total task score of 5.83 with a standard deviation of 2.74.

***Task 7***. Item 2 was the most difficult item, $p = 0.33$, while item 5 was the easiest, $p = 0.62$. Item 2 had the lowest standard deviation (0.94) while item 6 (field test) had the highest (1.00). Students averaged a total task score of 5.65 with a standard deviation of 2.91.

***Task 8.*** Item 2 was the most difficult item, $p = 0.37$, while item 3 was the easiest, $p = 0.66$. Item 3 had the lowest standard deviation (0.95) while item 4 had the highest (0.99). Students averaged a total task score of 5.20 with a standard deviation of 2.51.

***Task 9.*** Item 6 (field test) was the most difficult item, $p = 0.37$, while items 1 was the easiest, $p = 0.81$. Item 3 had the lowest standard deviation (0.78) while item 5 had the highest (0.97). Students averaged a total task score of 6.27 with a standard deviation of 2.46.

***Total test.*** The average total test score was 37.62 with a standard deviation of 20.87. Item difficulties range from $p$=.22 (the most difficult item) to $p$=.96 (the easiest item).

**Math: Grade 7**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties (*p* value) ranging from 0.87 – 0.95. Item 1 had the lowest standard deviation (0.63), while item 5 had the highest standard deviation (.97). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The range of items was much larger for task 2, and the items overall were more difficult. Items 5 and 6 were the most difficult item, $p = 0.31$, while item 1 was the easiest, $p = 0.58$. Items 5 and 6 had the lowest standard deviation (0.62) while item 3 had the highest (1.00). Students averaged a total task score of 5.37 with a standard deviation of 3.51.

***Task 3.*** Item 4 was the most difficult item, $p = 0.39$, while item 2 was the easiest, $p = 0.62$. Item 1 had the lowest standard deviation (0.82) while item 6 (field test) had the highest (1.00). Students averaged a total task score of 5.69 with a standard deviation of 3.41.

***Task 4.*** Item 6 (field test) was the most difficult item, $p = 0.46$, while items 3 and 5 were the easiest, $p = 0.75$. Item 3 had the lowest standard deviation (0.87) while item 1 had the highest (1.00). Students averaged a total task score of 7.18 with a standard deviation of 3.38.

***Task 5.*** Item 6 (field test) was the most difficult item, $p = 0.29$, while item 1 was the easiest, $p = 0.65$. Item 4 had the lowest standard deviation (0.65) while item 5 had the highest (1.00). Students averaged a total task score of 5.81 with a standard deviation of 2.89.

***Task 6.*** Item 4 was the most difficult item, $p = 0.37$, while item 2 was the easiest, $p = 0.82$. Item 5 had the lowest standard deviation (0.77) while item 6 (field test) had the highest (1.00). Students averaged a total task score of 7.44 with a standard deviation of 2.96.

***Task 7.*** Items 1 and 5 were the most difficult items, $p = 0.25$, while item 3 was the easiest, $p = 0.54$. Item 3 had the lowest standard deviation (0.72) while item 4 had the highest (1.00). Students averaged a total task score of 4.20 with a standard deviation of 2.48.

***Task 8.*** Item 2 was the most difficult item, $p = 0.48$, while item 3 was the easiest, $p = 0.64$. Item 3 had the lowest standard deviation (0.70) while item 4 had the highest (1.00). Students averaged a total task score of 5.70 with a standard deviation of 2.50.

***Task 9.*** Item 4 was the most difficult item, $p = 0.36$, while items 1 was the easiest, $p = 0.78$. Item 3 had the lowest standard deviation (0.83) while item 5 had the highest (0.98). Students averaged a total task score of 6.05 with a standard deviation of 2.60.

***Total test.*** The average total test score was 39.92 with a standard deviation of 23.20. Item difficulties range from $p=.25$ (the most difficult item) to $p=.95$ (the easiest item).

**Math: Grade 8**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties (*p* value) ranging from 0.87 – 0.95. Item 1 had the lowest standard deviation (0.63), while item 5 had the highest standard deviation (.97). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The range of items was much larger for task 2 and the items overall were more difficult. Item 4 was the most difficult item, $p = 0.26$, while item 1 was the easiest, $p = 0.58$. Item 4 had the lowest standard deviation (0.87) while item 3 had the highest (0.96). Students averaged a total task score of 5.37 with a standard deviation of 3.51.

***Task 3.*** Item 6 (field test) was the most difficult item, $p = 0.23$, while item 2 was the easiest, $p = 0.54$. Item 6 (field test) had the lowest standard deviation (0.84) while item 4 had the highest (1.00). Students averaged a total task score of 4.50 with a standard deviation of 3.05.

***Task 4.*** Item 5 was the most difficult item, $p = 0.27$, while item 3 was the easiest, $p = 0.64$. Item 5 had the lowest standard deviation (0.88) while item 1 had the highest (1.00). Students averaged a total task score of 5.84 with a standard deviation of 3.14.

***Task 5.*** Item 1 was the most difficult item, $p = 0.37$, while item 4 was the easiest, $p = 0.75$. Item 4 had the lowest standard deviation (0.87) while item 5 had the highest (1.00). Students averaged a total task score of 6.02 with a standard deviation of 3.32.

***Task 6.*** Item 1 was the most difficult item, $p = 0.64$, while item 3 was the easiest, $p = 0.82$. Item 3 had the lowest standard deviation (0.77) while item 1 had the highest (0.96). Students averaged a total task score of 7.14 with a standard deviation of 2.78.

***Task 7.*** Item 1 was the most difficult items, $p = 0.35$, while item 5 was the easiest, $p = 0.80$. Item 3 had the lowest standard deviation (0.75) while item 2 had the highest (0.98). Students averaged a total task score of 5.32 with a standard deviation of 2.57.

***Task 8.*** Item 5 was the most difficult item, $p = 0.35$, while item 3 was the easiest, $p = 0.74$. Item 3 had the lowest standard deviation (0.88) while item 2 had the highest (0.98). Students averaged a total task score of 5.42 with a standard deviation of 2.65.

***Task 9.*** Item 3 was the most difficult item, $p = 0.23$, while items 5 was the easiest, $p = 0.48$. Item 3 had the lowest standard deviation (0.85) while item 5 had the highest (1.00). Students averaged a total task score of 3.49 with a standard deviation of 2.44.

***Total test.*** The average total test score was 35.34 with a standard deviation of 21.25. Item difficulties range from $p$=.23 (the most difficult item) to $p$=.95 (the easiest item).

**Math: Grade 11**

***Task 1.*** were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.85 – 0.97. Item 1 had the lowest standard deviation (0.55), while item 5 had the highest standard deviation (1.05). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** the range was much larger, and the items overall were more difficult. Item 1 was the most difficult item, $p = 0.26$, while item 4 was the easiest, $p = 0.52$. Item 4 had the lowest standard deviation (0.77) while item 2 had the highest (0.99). Students averaged a total task score of 3.93 with a standard deviation of 2.85.

***Task 3.*** Item 4 was the most difficult item, $p = 0.24$, while item 2 was the easiest, $p = 0.50$. Item 4 had the lowest standard deviation (0.85) while item 2 had the highest (1.00). Students averaged a total task score of 3.54 with a standard deviation of 2.75.

***Task 4.*** Items 3 was the most difficult items, $p = 0.36$, while item 4 was the easiest, $p = 0.53$. Item 5 had the lowest standard deviation (0.80) while item 2 had the highest (1.00). Students averaged a total task score of 4.69 with a standard deviation of 2.90.

***Task 5.*** Item 4 was the most difficult item, $p = 0.24$, while item 5 was the easiest, $p = 0.52$. Item 5 had the lowest standard deviation (0.84) while item 2 had the highest (0.99). Students averaged a total task score of 3.77 with a standard deviation of 2.29.

***Task 6.*** Item 2 was the most difficult item, $p = 0.57$, while item 5 was the easiest, $p = 0.77$. Item 5 had the lowest standard deviation (0.74) while item 2 had the highest (0.99). Students averaged a total task score of 6.44 with a standard deviation of 2.95.

***Task 7.*** Item 2 was the most difficult items, $p = 0.28$, while item 5 was the easiest, $p = 0.54$. Item 2 had the lowest standard deviation (0.89) while item 4 had the highest (1.00). Students averaged a total task score of 4.50 with a standard deviation of 2.64.

***Task 8.*** Item 2 was the most difficult item, $p = 0.59$, while item 5 was the easiest, $p = 0.86$. Item 5 had the lowest standard deviation (0.71) while item 2 had the highest (0.98). Students averaged a total task score of 6.94 with a standard deviation of 2.85.

***Task 9.*** Item 3 was the most difficult item, $p = 0.46$, while items 2 was the easiest, $p = 0.88$. Item 2 had the lowest standard deviation (0.65) while item 3 had the highest (1.00). Students averaged a total task score of 5.93 with a standard deviation of 2.33.

***Total test.*** The average total test score was 35.34 with a standard deviation of 21.25. Item difficulties range from $p$=.24 (the most difficult item) to $p$=.97 (the easiest item).

**Science: Grade 5**

***Task 1***. Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.92 – 0.96. Generally, the more difficult items had a higher standard deviation than the less difficult items. For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** Item in task 2 had a much larger range and the items overall were more difficult. Item 4 was the most difficult item, $p = 0.60$, while item 2 was the easiest, $p = 0.84$. Item 2 had the lowest standard deviation (0.73) while item 4 had the highest (0.98). Students averaged a total task score of 7.18 with a standard deviation of 3.15.

***Task 3.*** Item 4 was the most difficult item, $p = 0.63$, while item 4 was the easiest, $p = 0.87$. Item 5 had the lowest standard deviation (0.67) while item 2 had the highest (0.90). Students averaged a total task score of 7.17 with a standard deviation of 3.29.

***Task 4.*** Item 5 was the most difficult item, $p = 0.57$, while items 3 was the easiest, $p = 0.80$. Item 3 had the lowest standard deviation (0.81) while items 5 was the highest (0.99). Students averaged a total task score of 6.66 with a standard deviation of 3.44.

***Task 5.*** Items 2 and 3 were the most difficult items, $p = 0.81$, while item 5 was the easiest, $p = 0.91$. Item 5 had the lowest standard deviation (0.57) while item 3 had the highest (0.79). Students averaged a total task score of 7.65 with a standard deviation of 3.44.

***Task 6.*** Item 5 was the most difficult item, $p = 0.75$, while item 1 was the easiest, $p = 0.93$. Item 1 had the lowest standard deviation (0.52) while item 5 had the highest (0.87). Students averaged a total task score of 7.55 with a standard deviation of 3.51.

***Task 7.*** Item 1 was the most difficult item, $p = 0.63$, while item 5 was the easiest, $p = 0.95$. Item 5 had the lowest standard deviation (0.46) while item 1 had the highest (0.97). Students averaged a total task score of 7.46 with a standard deviation of 3.43.

***Task 8.*** Item 4 was the most difficult item, $p = 0.39$, while item 2 was the easiest, $p = 0.88$. Item 2 had the lowest standard deviation (0.66) while item 5 had the highest (0.99). Students averaged a total task score of 5.80 with a standard deviation of 3.27.

***Task 9.*** Item 4 was the most difficult item, $p = 0.83$, while item 3 was the easiest, $p = 0.92$. Item 3 had the lowest standard deviation (0.56) while item 5 had the highest (0.73). Students averaged a total task score of 7.44 with a standard deviation of 3.77.

***Field test items.*** Item 1 in task 6 was the most difficult item, $p = 0.38$, while item 3 in task 8 was the easiest item, $p = 0.78$. Item 3 in task 8 had the lowest standard deviation (0.84), while item 1 in task 5 had the highest (1.01).

***Total test.*** The average total test score for the operational items was 64.98 with a standard deviation of 14.58. Item difficulties range from $p$=.38 (the most difficult item, a field test item) to $p$=.96 (the easiest item).

**Science: Grade 8**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.93 – 0.96. Item 1 had the lowest standard deviation (0.61), while item 3 had the highest standard deviation (.80). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The range of items was much larger for task 2 and the items overall were more difficult. Item 4 was the most difficult item, $p = 0.44$, while item 1 was the easiest, $p = 0.86$. Item 1 had the lowest standard deviation (0.71) while item 4 had the highest (0.99). Students averaged a total task score of 6.21 with a standard deviation of 3.07.

***Task 3.*** Item 1 was the most difficult item, $p = 0.52$, while item 3 was the easiest, $p = 0.84$. Item 3 had the lowest standard deviation (0.75) while item 1 had the highest (1.00). Students averaged a total task score of 6.91 with a standard deviation of 3.28.

***Task 4.*** Item 4 was the most difficult item, $p = 0.42$, while item 2 was the easiest, $p = 0.91$. Item 2 had the lowest standard deviation (0.57) while item 3 had the highest (1.00). Students averaged a total task score of 5.95 with a standard deviation of 3.32.

***Task 5.*** Item 1 was the most difficult item, $p = 0.22$, while item 2 was the easiest, $p = 0.87$. Item 2 had the lowest standard deviation (0.68) while items 3 and 4 had the highest (1.00). Students averaged a total task score of 4.98 with a standard deviation of 2.90.

***Task 6.*** Item 4 was the most difficult item, $p = 0.21$, while item 5 was the easiest, $p = 1.70$. Item 5 had the lowest standard deviation (0.72) while item 1 had the highest (0.99). Students averaged a total task score of 5.10 with a standard deviation of 3.22.

***Task 7.*** Item 2 was the most difficult item, $p = 0.56$, while item 5 was the easiest, $p = 0.80$. Item 5 had the lowest standard deviation (0.80) while item 2 had the highest (0.99). Students averaged a total task score of 5.85 with a standard deviation of 3.58.

***Task 8.*** Item 5 was the most difficult item, $p = 0.50$, while item 1 was the easiest, $p = 0.94$. Item 1 had the lowest standard deviation (0.48) while item 5 had the highest (1.00). Students averaged a total task score of 6.59 with a standard deviation of 3.46.

***Task 9.*** Item 2 was the most difficult item, $p = 0.60$, while items 1 and 3 was the easiest, $p = 0.83$. Item 3 had the lowest standard deviation (0.75) while item 2 had the highest (0.98). Students averaged a total task score of 6.32 with a standard deviation of 3.75.

***Field test items.*** in grade 8, item 3 in task 3 was the most difficult item, $p = 0.27$, while item 4 in task 7 was the easiest item, $p = 0.76$. Item 1 in task 7 had the lowest standard deviation (0.72), while item 1 in task 3 had the highest (1.01).

***Total test.*** The average total test score for the operational items was 56.20 with a standard deviation of 14.60. Item difficulties range from $p$=.21 (the most difficult item) to $p$=.96 (the easiest item).

**Science: Grade 11**

***Task 1.*** Items were scored on a 4-point scale. Task 1 items were quite easy, with item difficulties ($p$ value) ranging from 0.92 – 0.96. Item 1 had the lowest standard deviation (0.57), while item 9 had the highest standard deviation (.83). For Tasks 2-9 all items were scored on a 2-point scale.

***Task 2.*** The range of items was much larger for task 2 and the items overall were more difficult. Item 5 was the most difficult item, $p = 0.40$, while item 2 was the easiest, $p = 0.83$. Item 2 had the lowest standard deviation (0.75) while item 3 had the highest (0.99). Students averaged a total task score of 5.97 with a standard deviation of 3.15.

***Task 3.*** Item 5 was the most difficult item, $p = 0.48$, while item 1 was the easiest, $p = 0.73$. Item 1 had the lowest standard deviation (0.89) while item 4 had the highest (1.00). Students averaged a total task score of 5.94 with a standard deviation of 3.50.

***Task 4.*** Item 5 was the most difficult item, $p = 0.42$, while item 2 was the easiest, $p = 0.73$. Item 2 had the lowest standard deviation (0.88) while item 1 had the highest (1.00). Students averaged a total task score of 4.93 with a standard deviation of 3.43.

***Task 5.*** Item 3 was the most difficult item, $p = 0.38$, while item 2 was the easiest, $p = 0.82$. Item 2 had the lowest standard deviation (0.77) while item 3 had the highest (1.00). Students averaged a total task score of 5.41 with a standard deviation of 3.47.

***Task 6.*** Item 2 was the most difficult item, $p = 0.38$, while item 4 was the easiest, $p = 0.67$. Item 4 had the lowest standard deviation (0.94) while item 3 had the highest (1.00). Students averaged a total task score of 4.37 with a standard deviation of 3.25.

***Task 7.*** Item 1 was the most difficult item, $p = 0.34$, while item 5 was the easiest, $p = 0.81$. Item 1 had the lowest standard deviation (0.95) while item 2 had the highest (1.00). Students averaged a total task score of 4.70 with a standard deviation of 3.32.

***Task 8.*** Item 3 was the most difficult item, $p = 0.74$, while item 4 was the easiest, $p = 0.86$. Item 4 had the lowest standard deviation (0.71) while item 3 had the highest (0.89). Students averaged a total task score of 6.50 with a standard deviation of 3.97.

***Task 9.*** Item 4 was the most difficult item, $p = 0.42$, while items 5 was the easiest, $p = 0.72$. Item 5 had the lowest standard deviation (0.91) while item 1 had the highest (1.00). Students averaged a total task score of 4.58 with a standard deviation of 3.49.

***Field test items.*** in grade 11, item 4 in task 1 was the most difficult item, $p = 0.31$, while item 3 in task 8 was the easiest item, $p = 0.80$. Item 3 in task 8 had the lowest standard deviation (0.80), while item 2 in task 6 had the highest (1.01).

***Total test.*** The average total test score was 51.97 with a standard deviation of 15.79. Item difficulties range from $p$=.31 (the most difficult item, a field test item) to $p$=.96 (the easiest item).

## Analyses Within and Across Subject Areas

### Overview

The correlational analyses section below elaborates that the results we garnered were consistent with what we would expect, not too high and not too low (between .70 and .90). This indicates that we are indeed measuring different constructs (reading, writing, math, science). In Model 1, we explain that the pre-requisite skills task has only a small to moderate impact upon total test score, which is consistent with what we would expect given the test design where students are provided that level of support in order to mitigate factors that are construct-irrelevant throughout the remainder of the content prompts administered. In Model 2, we explain that the test type analysis results are also consistent with what we would expect, as the Scaffold version of the assessment is more difficult than the Standard version despite the provision of additional supports. This performance is attributed to the fact that lower-achieving students, in general, participate in the Scaffold assessment. Thus the test type differences appear to be student-related and not test-related. This is desirable. These results are within expected levels and do not demonstrate any threats to validity. In Model 3, we show that disability type and race/ethnicity have very little bearing upon the pre-requisite skills task results, which is what we would hope for on both accounts given the relatively homogeneous nature of this student population (who all have a significant cognitive disability). In Model 4, we conduct the same analysis as Model 3, but compared to total test scale score instead of pre-requisite skills task. Again, the results are within expected levels and do not demonstrate any threats to validity. In Model 5, we look at task functioning on the test. In other words, how well did the task separate student achievement compared to the total test (e.g., did students who performed poorly on the total test mostly miss this task and did students who performed well on the total test mostly get this task correct?). Again, the results are within expected ranges and no threats to validity are observed.

### Correlational Analyses

The correlations among students' total scores across subject areas were evaluated. The purpose of the analysis was to investigate how strongly a students score in one area "went along with" the students' scores in other subject areas. If the correlations were exceedingly high (e.g., above .90), it would indicate that the score a student receives in an individual subject has less to do with the intended construct (i.e., reading) than with factors idiosyncratic to the student. For example, if all subject areas correlated at .95, then it would provide strong evidence that the tests would be measuring a global student-specific construct (i.e., intelligence), and not the individual subject constructs. We would expect, however, that the tests would correlate quite strongly given that the same students were assessed multiple times. Therefore, we would expect moderately strong correlations (e.g., 0.7) simply because of the within-subject design. Idiosyncratic variance associated with the individual student is thus captured.

Full multi-trait mono-method analysis (correlation of content areas across grades) are reported in *Appendix D*. At grade 3, reading and math had a moderately strong correlation, $r(997) = .875$, $p < .05$. At grade 4, the correlation between reading and math ($n = 866$), reading and writing ($n = 863$), and math and writing ($n = 794$) were all moderately strong and statistically significant, with Pearson's $r$ in the .80's. At grade 5, reading and math had a moderately strong correlation, $r(860) = .858$, $p < .05$. The correlation between reading and science ($n = 743$) and math and science ($n = 739$) were moderate and significant, ranging in the .70's. At grade 5, reading and math had a moderately strong correlation, $r(690) = .828$, $p < .05$. At grade 7, correlation between reading and math ($n = 658$), reading and writing ($n = 651$), and math and writing ($n = 643$) were

all moderately strong and statistically significant, with Pearson's *r* in the .80's. At grade 8, correlation between reading and math ($n = 590$), reading and science ($n = 577$), and math and science ($n = 583$) were all moderately strong and statistically significant, with Pearson's *r* in the .80's. Finally, at grade 11, reading was statistically correlated to writing ($n = 315$), math ($n = 312$), and science ($n = 326$), with Pearson's *r* in the .80's. Math was statistically correlated to writing ($n = 315$) and science ($n = 326$), with Pearson's *r* in the .80's. Writing and science had a moderately strong correlation, $r(321) = .874$, $p < .05$.

**Regression Methods**

Several regression models were run to examine the functioning of the Oregon Alternate Assessment. Each model was run by grade-level for each subject: reading, writing, math, and science. These analyses provide information supporting the validity of inferences as a function of performance in a content subject area rather than pre-requisite skills, administration type, or disability categories. All continuous predictor variables were centered around the mean. Each model is outlined below:

*Regression Models*

| Model | Predictors | Dependent Variable |
|---|---|---|
| 1 | Pre-requisite task total | Score |
| 2 | Administration type<br>Pre-requisite task total | Score |
| 3 | Disability category<br>Administration type<br>Race/Ethnicity | Prerequisite Total Score |
| 4 | Disability category<br>Administration type<br>Race/Ethnicity | Score |
| 5 | Content Task totals<br>(separate models) | Score |

***Model 1***. Conducted to examine the relation between students pre-requisite skills task total with the total scale score for the test. In other words, did students scoring high on the pre-requisite skills task generally score high on the content tasks? The pre-requisite skills task assesses students' level of independence, while the total scale score assesses students' content knowledge. A strong relation between the pre-requisite skills task and the content tasks would indicate that the students' level of independence plays a large role in the content score they receive. Similarly, a low relation would indicate that students level of independence has very little to do with the

score they receive. It is important to note that the score the student received on the pre-requisite skills task also determines the level of support the student receives on the content tasks. Thus, we would expect the relation between the pre-requisite skills task total and the content task total to be quite low given that: (a) the tasks assess distinctly different constructs, and (b) students with lower levels of independence were supported during the content task administration to reduce the effect of any impeding factors that would preclude them from demonstrating their content knowledge. The full results are described on pages 6-8. Overall, the model accounted for between 17% - 57% of the total variance across subjects and grades.

***Model 2***. Conducted for a similar purpose as Model 1, but students' performance was also conditioned on the type of administration the student received: standard versus scaffold administration. Students taking the standard administration of the test were entered as the referent group. The scaffold administration has built in supports not available in the standard administration (i.e., auditory prompts by the Assessor). The extra supports are intended to minimize the effect of factors that would preclude students from demonstrating their content knowledge. However, it is also important to note that students taking the scaffold version of the test are generally lower performing students compared to those taking the standard version of the test. The type of administration a student receives is determined prior to the student taking the test by the student's IEP team. Thus, although the scaffold version helps students access the test and display their content knowledge, the observed effects cannot be attributed fully to the differences in test design. Rather, the observed effect represents the combined effects of the test design differences and the student group differences. The observed effects are generally quite large. Because students receiving the scaffold administration receive additional support, we would logically predict that the test would be easier than the standard administration. However, when inspecting the unstandardized regression weights (with standard administration as the referent group) it is apparent that students receiving the scaffold administration scored lower than students receiving the standard administration. Thus, the observed differences are likely due more to the student groups taking each version of the test than to the test itself. The full results are described on pages 9-14. Overall, the full model accounted for between 37% - 64% of the total variance across subjects and grades. In reading and writing test administration type generally accounted for the most unique variance, while in math and science pre-requisite skills task total accounted for the most unique variance.

***Model 3***. Conducted to examine how students with different disabilities and race/ethnicities performed on the pre-requisite skills task. Administration type was entered primarily as a control variable, but was also used to examine the proportion of students with each disability type in each administration type. Referent groups included students who were classified with a mental retardation disability, took the standard administration of the test, and were White. Each referent group was chosen based on the subgroup with the largest proportion of students. It was necessary to control for the variance associated with different administration types (scaffold versus standard), given that different student groups are represented in each (see results of Model 2). Holding administration constant, an examination of how students performed on the pre-requisite skills by the type of disability and race/ethnicity was performed. Hypothetically, the student's disability should play a role in the students' prerequisite skills score, given that the task is intended to assess students' level of independence. Different disability types could then logically be associated with different levels of independence. However, all students taking the assessment

also have a significant cognitive disability and we would therefore not expect disability to play a substantial role. Ideally, students' race/ethnicity would have essentially nothing to do with the score the student received on any portion of the test, including the prerequisite skills. The full results are described on pages 15-19. Overall, students' disability classification accounted for between 4% - 14% of the total variance across subjects and grades. Test administration type accounted for additional variance beyond students' disability (2% - 12%) and was generally the largest predictor. Students' race/ethnicity accounted for minimal variance when added in the third block (0% - 2%), and was generally not a statistically significant addition.

***Model 4.*** Conducted to examine how the same variables from Model 3 influenced students' total scale score. The reference groups were the same as in Model 3. Model 4 was conducted for the same purpose as Model 3, but to evaluate the effect of each variable on the students' content score (instead of prerequisite skills score). The full results for Model 4 are described on pages 20-24. Overall, students' disability classification accounted for between 13% - 27% of the total variance across subjects and grades. Test administration type accounted for additional variance beyond students' disability (10% - 24%) and was generally the largest predictor. Students' race/ethnicity accounted for minimal variance when added in the third block (0% - 2%), and was generally not a statistically significant addition.

***Model 5***. Conducted to examine the functioning of each task on each test. The degree to which each task did or did not predict performance on the total test was interpreted as a task-level discrimination index. For example, a task with a relatively low relation to the total scale score would indicate a more poorly functioning task than ones with a high relation. The logic followed that of a point-biserial correlation during an item-level analysis, but was applied at the task level. Refer to pages 25 – 34 for a full description of the relation between specific tasks and the scale score. Generally, all tasks had moderate to strong correlation to the total scale score, ranging from .39 to .90 across subjects and grades.

For all other models, reading was conducted by grade-level. For Model 5, all analyses were conducted at the grade-band level, given that the primary purpose of the analysis was to examine the function of each task, and not the characteristics or performance of students. For example, Model 1 was conducted to examine how *students* with different prerequisite task totals performed on the test. Thus, separation by grade-level was important to avoid a potential confound. The functioning of the pre-requisite task can also be examined with this analysis, but student performance was of primary concern. However, for Model 5 we were only interested in the functioning of the *tasks* and therefore conducted the analysis with the sample of students who took each task (i.e., grade-banded).

***Regression Procedures.*** Simultaneous regression was used for Model 2. Predictor variables were examined relative to the variables' regression weights (*b*) and unique contribution to the regression equation (semi-parial correlations).

Sequential regression was used for Models 3 and 4, with disability category entered into the first block, test administration type into second block, and race/ethnicity into the third block. Predictor variables were again examined relative to the variables' regression weights and unique

contribution to the regression equation. However, blocking variables into steps also allowed for an evaluation of the change in overall model fit between sets of variables.

***Assumptions.*** The Kolmogorov-Smirnov and Shapiro-Wilk tests of normality were used to examine the assumption of normality for all dependent variables. Frequency distributions and box-plots were also produced to visually interpret the assumption of normality. For all variables, the tests of normality were significant, indicating a non-normal distribution, and a visual examination of the frequency and box-plots confirmed non-normal distributions. However, the central limit theorem protects regression analyses from departures of normality as long as the sample size is reasonably large. Scatterplots were created for each predictor variable and corresponding dependent variable to examine the assumption of linearity. In all cases, the relation between the variables was roughly linear. Finally, the residuals were examined for normality with P-P and Q-Q plots, which revealed roughly normal distributions. The assumption of multicollinearity was investigated with the Tolerance and Variance Inflation Factor (VIF) statistics. All predictor variables were within the acceptable range, and are reported in the appendices for the respective models.

## Model 1 Results: Pre-req on Scale Scores

The full regression model, including correlations and descriptive statistics, are reported in *Appendix E*.

### Reading

***Grade 3.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 1145) = 365.71$, MSR = 94399.52, $p < .05$, = 0.24. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.01$, SE = .05, $p < .05$, 95% CI = .91 to 1.12. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.01 increase in students' scale scores.

***Grade 4.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 1001) = 208.93$, MSR = 59340.51, $p < .05$, = 0.17. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 0.87$, SE = .06, $p < .05$, 95% CI = 0.75 to 0.98. On average, every one point increase in the pre-requisite skills task total corresponded with a 0.87 increase in students' scale scores.

***Grade 5.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 973) = 212.36$, MSR = 69108.59, $p < .05$, = .18. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 0.89$, SE = .06, $p < .05$, 95% CI = 0.77 to 1.01. On average, every one point increase in the pre-requisite skills task total corresponded with a 0.89 increase in students' scale scores.

***Grade 6.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 747) = 278.76$, MSR = 91027.27, $p < .05$, = .27. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.17$, SE = .07, $p < .05$, 95% CI = 1.04 to 1.31. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.17 increase in students' scale scores.

***Grade 7.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 662) = 304.99$, MSR = 112894.57, $p < .05$, = .32. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.35$, SE = .08, $p < .05$, 95% CI = 1.20 to 1.51. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.35 increase in students' scale scores.

***Grade 8.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 601) = 214.00$, MSR = 82589.33, $p < .05$, = .26. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.19$, SE = .08, $p < .05$, 95% CI = 1.03 to 1.35. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.19 increase in students' scale scores.

***Grade 11.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 322) = 147.38$, MSR = 66343.42, $p < .05$, = .31. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.51$, SE = .13, $p < .05$, 95% CI = 1.29 to 1.78. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.51 increase in students' scale scores.

**Writing**

***Grade 4.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 887) = 215.44$, MSR = 96413.98, $p < .05$, = 0.20. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.16$, SE = .08, $p < .05$, 95% CI = 1.00 to 1.31. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.16 increase in students' scale scores.

***Grade 7.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 646) = 249.71$, MSR = 121986.19, $p < .05$, = .28. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.43$, SE = .09, $p < .05$, 95% CI = 1.25 to 1.61. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.43 increase in students' scale scores.

***Grade 11.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 303) = 168.28$, MSR = 72249.49, $p < .05$, = .36. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.54$, SE = .12, $p < .05$, 95% CI = 1.31 to 1.78. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.54 increase in students' scale scores.

**Math**

***Grade 3.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 780) = 1035.922$, MSR = 72196.669, $p < .05$, $R^2 = 0.57$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.362$, SE = .04, $p < .05$, 95% CI = 1.279 to1.445. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.362 increase in students' scale scores.

***Grade 4.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 677) = 655.159$, MSR = 74973.226, $p < .05$, $R^2 = 0.49$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.680$, SE = .07, $p < .05$, 95% CI = 1.551 to 1.809. On average, every one point increase in the pre-requisite skills task total corresponded with a 1. 680 increase in students' scale scores.

***Grade 5.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 699) = 865.347$, MSR = 77883.267, $p < .05$, $R^2 = 0.55$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.575$, SE = .05, $p < .05$, 95% CI = 1.469 to 1.575. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.575 increase in students' scale scores.

***Grade 6.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 564) = 571.710$, MSR = 36147.456, $p < .05$, $R^2 = 0.50$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.146$, SE = .05, $p < .05$, 95% CI = 1.051 to 1.240. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.146 increase in students' scale scores.

***Grade 7.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 558) = 669.232$, MSR = 53462.609, $p < .05$, $R^2 = 0.55$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.319$, SE = .05, $p < .05$, 95% CI = 1.219 to 1.419. On average, every one point increase in the pre-requisite skills task total corresponded with a 1. 319 increase in students' scale scores.

***Grade 8.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 487) = 353.297$, MSR = 53462.609, $p < .05$, $R^2 = 0.42$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.069$, SE = .06, $p < .05$,

95% CI = .957 to 1.181. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.069 in students' scale scores.

***Grade 11.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 270) = 267.312$, MSR = 36933.640, $p < .05$, $R^2 = 0.50$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.578$, SE = .10, $p < .05$, 95% CI = 1.388 to 1.768. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.578 in students' scale scores.

**Science**

***Grade 5.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 448) = 596.408$, MSR = 55454.203, $p < .05$, $R^2 = 0.48$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.626$, SE = .07, $p < .05$, 95% CI = 1.496 to 1.757. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.626 increase in students' scale scores.

***Grade 8.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 384) = 350.963$, MSR = 31265.503, $p < .05$, $R^2 = 0.42$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.348$, SE = .07, $p < .05$, 95% CI = 1.207 to 1.490. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.348 in students' scale scores.

***Grade 11.*** The regression of pre-requisite skills on students' scale score was statistically significant, $F(1, 210) = 213.551$, MSR = 15999.423, $p < .05$, $R^2 = 0.50$. Pre-requisite skills task total was a statistically significant predictor of students' scale score, $b = 1.338$, SE = .10, $p < .05$, 95% CI = 1.157 to 1.518. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.338 in students' scale scores.

**Model 2 Results (Simultaneous): Admin Type and Pre-req on Scale Scores**

The full regression model, including correlations and descriptive statistics, are reported in *Appendix F*.

**Reading**

***Grade 3.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 1144) = 390.60$, MSR = 79116.83, $p < .05$, = .41. Test administration type was a statistically significant predictor of students' scale score, $b$ = -16.25, SE = .92, $p < .05$, 95% CI = -18.05 to -14.46. On average, students taking the scaffold administration scored 16.25 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 0.78$, SE = .05, $p < .05$, 95% CI = 0.69 to 0.88. On average, every one point increase in the pre-requisite skills task total corresponded with a .78 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 16% of the total scale score variance was uniquely accounted for by test administration type, while 13% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 41% of the total variability in scale scores.

***Grade 4.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 1000) = 300.14$, MSR = 64452.96, $p < .05$, = .38. Test administration type was a statistically significant predictor of students' scale score, $b$ = -18.08, SE = 1.01, $p < .05$, 95% CI = -20.05 to 16.11. On average, students taking the scaffold administration scored 16.25 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 0.65$, SE = .05, $p < .05$, 95% CI = 0.55 to 0.76. On average, every one point increase in the pre-requisite skills task total corresponded with a .65 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 20% of the total scale score variance was uniquely accounted for by test administration type, while 9% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 38% of the total variability in scale scores.

***Grade 5.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 972) = 346.81$, MSR = 80321.42, $p < .05$, = .42. Test administration type was a statistically significant predictor of students' scale score, $b$ = -20.78, SE = 1.05, $p < .05$, 95% CI = -22.83 to -18.72. On average, students taking the scaffold administration scored 20.78 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 0.66$, SE = .05, $p < .05$, 95% CI = 0.56 to 0.77. On average, every one point increase in the pre-requisite skills task total corresponded with a .66 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 24% of the total scale score variance was uniquely accounted for by test administration type, while 9% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 42% of the total variability in scale scores.

***Grade 6.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 746) = 346.78$, MSR = 80689.17, $p < .05$, = .48. Test administration type was a statistically significant predictor of students' scale score, $b$ = -21.13, SE = 1.22, $p < .05$, 95% CI = -23.51 to -18.74. On average, students taking the scaffold administration scored 21.13 scale score points lower than students taking the standard

administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 0.84$, SE = .06, $p < .05$, 95% CI = 0.72 to 0.96. On average, every one point increase in the pre-requisite skills task total corresponded with a .78 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 21% of the total scale score variance was uniquely accounted for by test administration type, while 13% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 48% of the total variability in scale scores.

***Grade 7.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 661) = 336.69$, MSR = 90314.90, $p < .05$, = .51. Test administration type was a statistically significant predictor of students' scale score, $b = -21.31$, SE = 1.34, $p < .05$, 95% CI = -23.94 to -18.67. On average, students taking the scaffold administration scored 21.31 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.01$, SE = .07, $p < .05$, 95% CI = 0.88 to 1.15. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.01 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 19% of the total scale score variance was uniquely accounted for by test administration type, while 16% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 51% of the total variability in scores.

***Grade 8.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 600) = 267.77$, MSR = 74168.11, $p < .05$, = .47. Test administration type was a statistically significant predictor of students' scale score, $b = -22.00$, SE = .1.43, $p < .05$, 95% CI = -24.80 to -19.19. On average, students taking the scaffold administration scored 22.00 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 0.87$, SE = .07, $p < .05$, 95% CI = 0.73 to 1.02. On average, every one point increase in the pre-requisite skills task total corresponded with a 0.87 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 21% of the total scale score variance was uniquely accounted for by test administration type, while 14% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 47% of the total variability in scale scores.

***Grade 11.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 321) = 138.74$, MSR = 48982.27, $p < .05$, = .46. Test administration type was a statistically significant predictor of students' scale score, $b = -20.88$, SE = 2.21, $p < .05$, 95% CI = -25.22 to -16.54. On average, students taking the scaffold administration scored 20.88 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.20$, SE = .12, $p < .05$, 95% CI = 0.97 to 1.42. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.20 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 15% of the total scale score variance was uniquely accounted for by test administration type, while 18% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 46% of the total variability in scale scores.

**Writing**

***Grade 4.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 886) = 256.17$, MSR = 90383.17, $p < .05$, = .37. Test administration type was a statistically significant predictor of students' scale score, $b = -21.07$, SE = 1.36, $p < .05$, 95% CI = -23.74 to -18.39. On average, students taking the scaffold administration scored 21.07 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 0.87$, SE = .05, $p < .07$, 95% CI = 0.73 to 1.01. On average, every one point increase in the pre-requisite skills task total corresponded with a .87 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 17% of the total scale score variance was uniquely accounted for by test administration type, while 10% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 37% of the total variability in scale scores.

***Grade 7.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 645) = 291.66$, MSR = 103898.96, $p < .05$, = .48. Test administration type was a statistically significant predictor of students' scale score, $b = -24.41$, SE = 1.57, $p < .05$, 95% CI = -27.49 to -21.32. On average, students taking the scaffold administration scored 24.41 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.02$, SE = .08, $p < .05$, 95% CI = 0.86 to 1.18. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.02 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 20% of the total scale score variance was uniquely accounted for by test administration type, while 13% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 48% of the total variability in scale scores.

***Grade 11.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 302) = 142.15$, MSR = 49057.49, $p < .05$, = .49. Test administration type was a statistically significant predictor of students' scale score, $b = -20.50$, SE = 2.36, $p < .05$, 95% CI = -25.10 to -15.80. On average, students taking the scaffold administration scored 20.50 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.15$, SE = .12, $p < .05$, 95% CI = 0.92 to 1.37. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.15 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 13% of the total scale score variance was uniquely accounted for by test administration type, while 17% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 49% of the total variability in scale scores.

**Math**

***Grade 3.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 779) = 654.224$, MSR = 39664.136, $p < .05$, $R^2$ = .63. Test administration type was a statistically significant predictor of students' scale score, $b = -7.035$, SE = .65, $p < .05$, 95% CI = -8.308 to -5.762. On average, students taking the scaffold administration scored 7.035 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.12$, SE = .05, $p < .05$, 95% CI = 1.03 to 1.21. On average, every one

point increase in the pre-requisite skills task total corresponded with a 1.12 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 6% of the total scale score variance was uniquely accounted for by test administration type, while 30% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 63% of the total variability in scale scores.

***Grade 4.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 676) = 412.381$, MSR = 41889.207, $p < .05$, $R^2 = .55$. Test administration type was a statistically significant predictor of students' scale score, $b = -8.244$, SE = .89, $p < .05$, 95% CI = -9.983 to -6.506. On average, students taking the scaffold administration scored 8.244 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.390$, SE = .07, $p < .05$, 95% CI = 1.254 to 1.526. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.39 increase in students' scores. Examination of the squared semipartial correlations revealed that 6% of the total scale score variance was uniquely accounted for by test administration type, while 27% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 55% of the total variability in scale scores.

***Grade 5.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 698) = 589.565$, MSR = 44220.581, $p < .05$, $R^2 = .63$. Test administration type was a statistically significant predictor of students' scale score, $b = -8.883$, SE = .75, $p < .05$, 95% CI = -10.353 to -7.413. On average, students taking the scaffold administration scored 8.883 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.277$, SE = .06, $p < .05$, 95% CI = 1.169 to 1.385. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.277 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 8% of the total scale score variance was uniquely accounted for by test administration type, while 29% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 63% of the total variability in scale scores.

***Grade 6.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 563) = 352.021$, MSR = 19950.204, $p < .05$, $R^2 = .56$. Test administration type was a statistically significant predictor of students' scale score, $b = -5.987$, SE = .74, $p < .05$, 95% CI = -7.432 to -4.542. On average, students taking the scaffold administration scored 8.883 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = .942$, SE = .05, $p < .05$, 95% CI = .840 to 1.044. On average, every one point increase in the pre-requisite skills task total corresponded with a .942 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 5% of the total scale score variance was uniquely accounted for by test administration type, while 26% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 56% of the total variability in scale scores.

***Grade 7.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 557) = 444.146$, MSR = 30128.018, $p < .05$, $R^2 = .62$. Test administration type was a statistically significant predictor of students' scale score, $b = -7.917$, SE = .79, $p < .05$, 95% CI = -9.471 to -6.363. On average, students taking the scaffold administration scored 7.917 scale score points lower than students taking the standard

administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.069$, SE = .05, $p < .05$, 95% CI = .965 to 1.174. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.069 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 7% of the total scale score variance was uniquely accounted for by test administration type, while 28% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 62% of the total variability in scale scores.

***Grade 8.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 486) = 234.572$, MSR = 17072.943, $p < .05$, $R^2 = .49$. Test administration type was a statistically significant predictor of students' scale score, $b = -7.048$, SE = .86, $p < .05$, 95% CI = -8.732 to -5.363. On average, students taking the scaffold administration scored 7.048 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = .862$, SE = .06, $p < .05$, 95% CI = .746 to .978. On average, every one point increase in the pre-requisite skills task total corresponded with a .862 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 7% of the total scale score variance was uniquely accounted for by test administration type, while 22% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 49% of the total variability in scale scores.

***Grade 11.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 269) = 211.879$, MSR = 22705.768, $p < .05$, $R^2 = .61$. Test administration type was a statistically significant predictor of students' scale score, $b = -12.241$, SE = 1.38, $p < .05$, 95% CI = -14.951 to -9.532. On average, students taking the scaffold administration scored 12.241 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.249$, SE = .09, $p < .05$, 95% CI = 1.067 to 1.431. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.249 increase in students' scores. Examination of the squared semipartial correlations revealed that 11% of the total scale score variance was uniquely accounted for by test administration type, while 26% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 61% of the total variability in scale scores.

**Science**

***Grade 5.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 447) = 401.040$, MSR = 31178.749, $p < .05$, $R^2 = .64$. Test administration type was a statistically significant predictor of students' scale score, $b = -8.619$, SE = .915, $p < .05$, 95% CI = -10.417 to -6.822. On average, students taking the scaffold administration scored 8.619 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.372$, SE = .07, $p < .05$, 95% CI = 1.241 to 1.503. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.372 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 7% of the total scale score variance was uniquely accounted for by test administration type, while 34% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 64% of the total variability in scale scores.

***Grade 8.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 383) = 261.491$, MSR = 18897.606, $p < .05$, $R^2 = .58$. Test administration type was a statistically significant predictor of students' scale score, $b = -8.821$, SE = .928, $p < .05$, 95% CI = -10.645 to -6.996. On average, students taking the scaffold administration scored 8. 821 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.118$, SE = .07, $p < .05$, 95% CI = .982 to 1.254. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.118 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 10% of the total scale score variance was uniquely accounted for by test administration type, while 29% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 58% of the total variability in scale scores.

***Grade 11.*** The regression of pre-requisite skills and test administration type on students' scale score was statistically significant, $F(2, 209) = 165.820$, MSR = 18897.606, $p < .05$, $R^2 = .61$. Test administration type was a statistically significant predictor of students' scale score, $b = -8.645$, SE = 1.125, $p < .05$, 95% CI = -10.863 to -6.428. On average, students taking the scaffold administration scored 8. 645 scale score points lower than students taking the standard administration. Students' pre-requisite task total was also a statistically significant predictor of students' scale score, $b = 1.105$, SE = .09, $p < .05$, 95% CI = .934 to 1.275. On average, every one point increase in the pre-requisite skills task total corresponded with a 1.105 increase in students' scale scores. Examination of the squared semipartial correlations revealed that 11% of the total scale score variance was uniquely accounted for by test administration type, while 30% was uniquely accounted for by the pre-requisite task total. Together test administration type and Pre-requisite skills accounted for 61% of the total variability in scale scores.

**Model 3 Results (Sequential): Dis, Admin, & Race/Ethnicity on Pre-Req**

The full regression model, including correlations and descriptive statistics, are reported in *Appendix G*.

**Reading**

***Grade 3.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 1075) = 9.89$, MSR = 715.80, $p < .05$, $R^2 = .08$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 1074) = 37.83, MSR = 893.29, $p < .05$, $R^2$ *Change* = .03. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 1067) = 1.06, MSR = 574.12, $p = .39$, $R^2$ *Change* = .01. For the final model, deaf-blindness had the largest regression weight, $b = -28.11$, SE = 8.39, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 3% of the total variance in students Pre-Requisite skills total.

***Grade 4.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 956) = 7.23$, MSR = 542.15, $p < .05$, $R^2 = .07$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 955) = 26.07, MSR = 666.08, $p < .05$, $R^2$ *Change* = .03. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 948) = 0.50, MSR = 421.21, $p = .84$, $R^2$ *Change* = .00. For the final model, orthopedic impairment had the largest regression weight, $b = -10.42$, SE = 1.86, $p < .05$, and accounted for the most variance, uniquely accounting for 3% of the total variance in students Pre-Requisite skills total.

***Grade 5.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 925) = 4.09$, MSR = 354.56, $p < .05$, $R^2 = .04$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 924) = 27.75, MSR = 551.81, $p < .05$, $R^2$ *Change* = .03. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 917) = 0.94, MSR = 357.71, $p = .48$, $R^2$ *Change* = .01. For the final model, traumatic brain injury had the largest regression weight, $b = -12.34$, SE = 3.81, $p < .05$, and accounted for the most variance, uniquely accounting for 3% of the total variance in students Pre-Requisite skills total.

***Grade 6.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 722) = 6.40$, MSR = 528.42, $p < .05$, $R^2 = .07$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 721) = 50.79, MSR = 867.98, $p < .05$, $R^2$ *Change* = .06. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 714) = 1.75, MSR = 565.80, $p = .10$, $R^2$ *Change* = .02. For the final model, Asian/Pacific Islander had the largest regression weight, $b = 12.33$, SE = 6.14, $p < .05$, while administration type accounted for the most variance, uniquely accounting for 6% of the total variance in students Pre-Requisite skills total.

***Grade 7.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 633) = 5.18$, MSR = 460.24, $p < .05$, $R^2 = .07$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 632) = 47.28, MSR = 805.60, $p < .05$, $R^2$ *Change* = .07. For the third block, students race/ethnicity was added to the model, which result in a significant change in model fit, *F Change*(7, 625) = 2.51, MSR = 557.89, $p < .05$, $R^2$ *Change* =

.02. For the final model, Asian/Pacific Islander had the largest regression weight, $b = 16.93$, SE = 11.08, $p = .13$, while administration type accounted for the most variance, uniquely accounting for 7% of the total variance in students Pre-Requisite skills total.

***Grade 8.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 581) = 8.20$, MSR = 707.93, $p < .05$, $R^2 = .11$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*$(1, 580) = 25.03$, MSR = 844.79, $p < .05$, $R^2$ *Change* = .04. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*$(7, 573) = 1.73$, MSR = 555.56, $p = .10$, $R^2$ *Change* = .02. For the final model, visual impairment had the largest regression weight, $b = -15.36$, SE = 3.55, $p < .05$, while administration type accounted for the most variance, uniquely accounting for 4% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 297) = 3.17$, MSR = 260.99, $p < .05$, $R^2 = .09$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*$(1, 296) = 21.28$, MSR = 398.70, $p < .05$, $R^2$ *Change* = .06. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*$(7, 289) = 0.26$, MSR = 242.89, $p = .97$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -16.79$, SE = 6.33, $p < .05$, while administration type accounted for the most variance, uniquely accounting for 6% of the total variance in students Pre-Requisite skills total.

**Writing**

***Grade 4.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 851) = 9.01$, MSR = 662.52, $p < .05$, $R^2 = .10$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*$(1, 850) = 30.24$, MSR = 797.78, $p < .05$, $R^2$ *Change* = .03. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*$(7, 843) = 0.19$, MSR = 492.76, $p = .99$, $R^2$ *Change* = .00. For the final model, deaf-blindness had the largest regression weight, $b = -16.19$, SE = 8.49, $p = .06$, while test administration type accounted for the most variance, uniquely accounting for 3% of the total variance in students Pre-Requisite skills total.

***Grade 7.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 614) = 3.73$, MSR = 339.38, $p < .05$, $R^2 = .05$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*$(1, 613) = 54.03$, MSR = 757.48, $p < .05$, $R^2$ *Change* = .08. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*$(6, 607) = 1.44$, MSR = 518.27, $p = .20$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -6.67$, SE = 4.61, $p = .15$, while test administration type accounted for the most variance, uniquely accounting for 8% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 79) = 3.65$, MSR = 338.30, $p < .05$, $R^2 = .11$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*$(1, 278) = 44.66$, MSR = 662.85, $p < .05$, $R^2$ *Change* = .12. For the third block, students race/ethnicity was added to the model, which did not result in a significant

change in model fit, *F Change*(7, 271) = 0.20, MSR = 396.74, $p = .99$, $R^2$ *Change* = .00. For the final model, visual impairment had the largest regression weight, $b = -15.03$, SE = 6.47, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 12% of the total variance in students Pre-Requisite skills total.

**Math**

***Grade 3.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 931) = 7.02$, MSR = 534.76, $p < .05$, $R^2 = .07$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 930) = 46.93, MSR = 796.03, $p < .05$, $R^2$ *Change* = .05. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 923) = 0.46, MSR = 499.60, $p = .86$, $R^2$ *Change* = .00. For the final model, deaf-blindness had the largest regression weight, $b = -24.66$, SE = 8.57, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 4% of the total variance in students Pre-Requisite skills total.

***Grade 4.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 831) = 9.28$, MSR = 623.80, $p < .05$, $R^2 = .10$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 830) = 21.96, MSR = 698.04, $p < .05$, $R^2$ *Change* = .02. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 823) = 0.65, MSR = 698.04, $p = .65$, $R^2$ *Change* = .01. For the final model, deaf-blindness had the largest regression weight, $b = -17.33$, SE = 8.14, $p < .05$, while orthopedic impairment accounted for the most variance, uniquely accounting for 4% of the total variance in students Pre-Requisite skills total.

***Grade 5.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 839) = 4.15$, MSR = 331.02, $p < .05$, $R^2 = .04$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 838) = 20.61, MSR = 458.43, $p < .05$, $R^2$ *Change* = .02. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 831) = 0.53, MSR = 286.58, $p = .82$, $R^2$ *Change* = .00. For the final model, traumatic brain injury had the largest regression weight, $b = -8.74$, SE = 3.67, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 2% of the total variance in students Pre-Requisite skills total.

***Grade 6.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 688) = 5.22$, MSR = 468.22, $p < .05$, $R^2 = .06$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 687) = 40.70, MSR = 766.49, $p < .05$, $R^2$ *Change* = .05. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 680) = 1.18, MSR = 491.82, $p = .32$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -9.78$, SE = 4.65, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 5% of the total variance in students Pre-Requisite skills total.

***Grade 7.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 631) = 4.81$, MSR = 411.87, $p < .05$, $R^2 = .06$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 630) = 55.85, MSR = 810.90, $p < .05$, $R^2$ *Change* =

.08. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 624) = 1.30, MSR = 545.08, $p = .26$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -10.20$, SE = 4.01, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 8% of the total variance in students Pre-Requisite skills total.

***Grade 8.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 571) = 6.71$, MSR = 602.49, $p < .05$, $R^2 = .10$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 570) = 21.71, MSR = 730.49, $p < .05$, $R^2$ *Change* = .08. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 563) = 1.39, MSR = 479.17, $p = .21$, $R^2$ *Change* = .02. For the final model, visual impairment had the largest regression weight, $b = -17.97$, SE = 3.64, $p < .05$, and accounted for the most variance, uniquely accounting for 4% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 300) = 4.57$, MSR = 367.47, $p < .05$, $R^2 = .12$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 299) = 30.22, MSR = 552.19, $p < .05$, $R^2$ *Change* = .08. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 292) = 0.82, MSR = 349.63, $p = .57$, $R^2$ *Change* = .02. For the final model, visual impairment had the largest regression weight, $b = -18.11$, SE = 6.13, $p < .05$, and accounted for the most variance, uniquely accounting for 8% of the total variance in students Pre-Requisite skills total.

**Science**

***Grade 5.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 514) = 4.10$, MSR = 288.20, $p < .05$, $R^2 = .07$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 513) = 26.86, MSR = 288.20, $p < .05$, $R^2$ *Change* = .05. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 506) = 0.31, MSR = 266.98, $p = .95$, $R^2$ *Change* = .00. For the final model, traumatic brain injury had the largest regression weight, $b = -10.25$, SE = 3.76, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 5% of the total variance in students Pre-Requisite skills total.

***Grade 8.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 439) = 5.54$, MSR = 460.12, $p < .05$, $R^2 = .10$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 438) = 6.59, MSR = 468.18, $p < .05$, $R^2$ *Change* = .01. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 431) = 0.97, MSR = 308.04, $p = .46$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -24.63$, SE = 4.72, $p < .05$, and accounted for the most variance, uniquely accounting for 6% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 234) = 4.13$, MSR = 245.15, $p < .05$, $R^2 = .14$. Test Administration Type was added to the model for the second block, which resulted in a

significant change in model fit, *F Change*(1, 233) = 20.51, MSR = 333.10, $p < .05$, $R^2$ *Change* = .07. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 226) = 0.36, MSR = 204.18, $p = .93$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -15.16$, SE = 5.36, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 7% of the total variance in students Pre-Requisite skills total.

**Model 4 Results (Sequential): Dis, Admin, & Race/Ethnicity on Scale Score**

The full regression model, including correlations and descriptive statistics, are reported in *Appendix H*.

**Reading**

***Grade 3.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 1151) = 41.02$, MSR = 10032.98, $p < .05$, $R^2 = .26$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 1150) = 184.74, MSR = 12662.91, $p < .05$, $R^2$ *Change* = .10. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 1143) = 0.31, MSR = 7763.64, $p = .95$, $R^2$ *Change* = .00. For the final model, deaf-blindness had the largest regression weight, $b = -34.45$, SE = 14.60, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 10% of the total variance in students Pre-Requisite skills total.

***Grade 4.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 1042) = 34.52$, MSR = 8436.18, $p < .05$, $R^2 = .25$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 1041) = 218.36, MSR = 7189.34, $p < .05$, $R^2$ *Change* = .13. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 1034) = 0.63, MSR = 7189.34, $p = .73$, $R^2$ *Change* = .00. For the final model, visual impairment had the largest regression weight, $b = -18.29$, SE = 5.91, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 13% of the total variance in students Pre-Requisite skills total.

***Grade 5.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 1007) = 41.18$, MSR = 11731.94, $p < .05$, $R^2 = .27$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1,1006) = 207.48, MSR = 15463.88, $p < .05$, $R^2$ *Change* = .13. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 999) = 1.20, MSR = 9212.76, $p = .30$, $R^2$ *Change* = .01. For the final model, administration type had the largest regression weight, $b = -16.98$, SE = 1.19, $p < .05$, and accounted for the most variance, uniquely accounting for 12% of the total variance in students Pre-Requisite skills total.

***Grade 6.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 764) = 28.05$, MSR = 10078.67, $p < .05$, $R^2 = .25$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 763) = 205.59, MSR = 14898.57, $p < .05$, $R^2$ *Change* = .16. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 756) = 1.18, MSR = 8901.15, $p = .31$, $R^2$ *Change* = .01. For the final model, administration type was the largest regression weight, $b = -20.30$, SE = 1.43, $p < .05$, and accounted for the most variance, uniquely accounting for 21% of the total variance in students Pre-Requisite skills total.

***Grade 7.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 699) = 19.12$, MSR = 8811.40, $p < .05$, $R^2 = .20$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 698) = 230.26, MSR = 15919.49, $p < .05$, $R^2$ *Change* = .20. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 691) = 0.63, MSR = 9508.82, $p = .42$, $R^2$

*Change* = .01. For the final model, visual impairment had the largest regression weight, $b$ = -32.60, SE = 7.15, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 20% of the total variance in students Pre-Requisite skills total.

***Grade 8.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 630) = 20.16$, MSR = 8599.92, $p < .05$, $R^2 = .22$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 629) = 188.03, MSR = 13924.60, $p < .05$, $R^2$ *Change* = .18. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 622) = 1.00, MSR = 8326.13, $p = .43$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b$ = -26.24, SE = 7.09, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 18% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 318) = 6.20$, MSR = 3612.97, $p < .05$, $R^2 = .15$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 317) = 86.36, MSR = 7217.27, $p < .05$, $R^2$ *Change* = .18. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 310) = 0.90, MSR = 4415.23, $p = .51$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b$ = -32.62, SE = 10.97, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 18% of the total variance in students Pre-Requisite skills total.

**Writing**

***Grade 4.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 934) = 32.96$, MSR = 12900.12, $p < .05$, $R^2 = .26$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 933) = 156.71, MSR = 16506.53, $p < .05$, $R^2$ *Change* = .11. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 926) = 0.88, MSR = 10201.69, $p = .53$, $R^2$ *Change* = .00. For the final model, specific learning disability had the largest regression weight, $b$ = 27.68, SE = 1.96, $p < .05$, and accounted for the most variance, uniquely accounting for 16% of the total variance in students Pre-Requisite skills total.

***Grade 7.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 678) = 13.59$, MSR = 8064.80, $p < .05$, $R^2 = .15$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 677) = 267.90, MSR = 18664.96, $p < .05$, $R^2$ *Change* = .24. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 671) = 0.84, MSR = 11799.67, $p = .54$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b$ = -31.61, SE = 7.93, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 24% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 308) = 5.89$, MSR = 3338.01, $p < .05$, $R^2 = .15$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 307) = 106.85, MSR = 7512.25, $p < .05$, $R^2$ *Change* = .22. For the third block, students race/ethnicity was added to the model, which did not result in

a significant change in model fit, *F Change*(7, 300) = 1.11, MSR = 4610.66, $p = .36$, $R^2$ *Change* = .02. For the final model, test administration type had the largest regression weight, $b = -26.86$, SE = 2.60, $p < .05$, and accounted for the most variance, uniquely accounting for 22% of the total variance in students Pre-Requisite skills total.

**Math**

***Grade 3.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 992) = 30.62$, MSR = 3552.60, $p < .05$, $R^2 = .24$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 991) = 238.51, MSR = 5259.14, $p < .05$, $R^2$ *Change* = .15. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 984) = 0.66, MSR = 3237.83, $p = .71$, $R^2$ *Change* = .00. For the final model, deaf-blindness had the largest regression weight, $b = -22.47$, SE = 9.72, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 15% of the total variance in students Pre-Requisite skills total.

***Grade 4.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(10, 896) = 31.72$, MSR = 5070.81, $p < .05$, $R^2 = .26$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 895) = 217.25, MSR = 7152.92, $p < .05$, $R^2$ *Change* = .14. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 888) = 0.85, MSR = 4413.59, $p = .55$, $R^2$ *Change* = .00. For the final model, traumatic brain injury had the largest regression weight, $b = 13.68$, SE = 6.61, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 14% of the total variance in students Pre-Requisite skills total.

***Grade 5.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 906) = 27.20$, MSR = 3947.05, $p < .05$, $R^2 = .21$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 905) = 240.84, MSR = 6316.10, $p < .05$, $R^2$ *Change* = .17. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 898) = 1.42, MSR = 3782.06, $p = .20$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -18.58$, SE = 6.24, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 16% of the total variance in students Pre-Requisite skills total.

***Grade 6.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 731) = 22.93$, MSR = 2139.22, $p < .05$, $R^2 = .22$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 730) = 192.26, MSR = 3347.08, $p < .05$, $R^2$ *Change* = .16. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 723) = 1.68, MSR = 2019.70, $p = .11$, $R^2$ *Change* = .01. For the final model, Asian/Pacific Islander had the largest regression weight, $b = 11.10$, SE = 6.03, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 16% of the total variance in students Pre-Requisite skills total.

***Grade 7.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 695) = 15.71$, MSR = 2345.15, $p < .05$, $R^2 = .17$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 694) = 232.96, MSR = 4718.81, $p < .05$, $R^2$ *Change*

= .21. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 688) = 1.62, MSR = 3016.72, $p = .14$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -17.31$, SE = 4.05, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 22% of the total variance in students Pre-Requisite skills total.

***Grade 8.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 615) = 18.13$, MSR = 2054.39, $p < .05$, $R^2 = .21$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 614) = 126.44, MSR = 3039.28, $p < .05$, $R^2$ *Change* = .14. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 607) = 1.00, MSR = 1826.58, $p = .43$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -16.73$, SE = 3.79, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 13% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 325) = 5.36$, MSR = 1286.95, $p < .05$, $R^2 = .13$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 324) = 95.71, MSR = 2937.34, $p < .05$, $R^2$ *Change* = .20. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 317) = 0.56, MSR = 1770.96, $p = .79$, $R^2$ *Change* = .01. For the final model, Asian had the largest regression weight, $b = -20.10$, SE = 14.43, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 19% of the total variance in students Pre-Requisite skills total.

**Science**

***Grade 5.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 724) = 15.08$, MSR = 2916.24, $p < .05$, $R^2 = .16$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 723) = 137.98, MSR = 4868.25, $p < .05$, $R^2$ *Change* = .14. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 716) = 1.33, MSR = 2952.11, $p = .24$, $R^2$ *Change* = .01. For the final model, traumatic brain injury had the largest regression weight, $b = -13.52$, SE = 5.79, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 13% of the total variance in students Pre-Requisite skills total.

***Grade 8.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 599) = 15.89$, MSR = 2256.07, $p < .05$, $R^2 = .19$. Test Administration Type was added to the model for the second block, which resulted in a significant change in model fit, *F Change*(1, 598) = 156.87, MSR = 3797.38, $p < .05$, $R^2$ *Change* = .17. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 591) = 1.21, MSR = 2289.51, $p = .30$, $R^2$ *Change* = .01. For the final model, visual impairment had the largest regression weight, $b = -19.17$, SE = 4.15, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 17% of the total variance in students Pre-Requisite skills total.

***Grade 11.*** The first block of the regression model included only students' disability category, and was statistically significant, $F(9, 366) = 5.99$, MSR = 893.62, $p < .05$, $R^2 = .13$. Test Administration Type was added to the model for the second block, which resulted in a

significant change in model fit, *F Change*(1, 365) = 122.93, MSR = 2179.35, $p < .05$, $R^2$ *Change* = .22. For the third block, students race/ethnicity was added to the model, which did not result in a significant change in model fit, *F Change*(7, 358) = 0.80, MSR = 1318.77, $p = .59$, $R^2$ *Change* = .01. For the final model, Asian/Pacific Islander had the largest regression weight, $b = 15.65$, SE = 10.45, $p < .05$, while test administration type accounted for the most variance, uniquely accounting for 21% of the total variance in students Pre-Requisite skills total.

## Model 5 Results: Content Task Totals on Scale Score

The full regression model, including correlations and descriptive statistics, are reported in *Appendix I*.

### Reading

For reading there were 10 content tasks and thus 10 separate regression models for each grade. Below is a summary of all models within each grade-band.

***Elementary.*** The regression of each content task total score on students' scale score was statistically significant for each task. The first task assessing students content knowledge (Task 2) had the lowest correlation with students' total scale score, $r = .78$. Task 6 had the strongest correlation with students total scale score, $r = .87$. Overall, every one point gain in students' task scores corresponded to a 4.4 (Task 2) to 5.2 (Task 3) point gain in students' scale score.

***Middle.*** The regression of each content task total score on students' scale score was statistically significant for each task. Task 3 had the lowest correlation with students' total scale score, $r = .84$. Task 8 had the strongest correlation with students' total scale score, $r = .90$. Overall, every one point gain in students' task scores corresponded to a 4.9 (Task 2) to 6.1 (Task 2) point gain in students' scale score.

***High School.*** The regression of each content task total score on students' scale score was statistically significant for each task. The first task assessing students' content knowledge (Task 2) had the lowest correlation with students' total scale score, $r = .86$. Task 6 had the strongest correlation with students' total scale score, $r = .89$. Overall, every one point gain in students' task scores corresponded to a 5.5 (Task 3) to 6.8 (Task 2) point gain in students' scale score.

### Writing

For writing there were 10 content tasks and thus 10 separate regression models for each grade. Below is a summary of all models within one grade.

***Grade 4.*** The regression of each content task total score on students' scale score was statistically significant for each task. Task 10 had the lowest correlation with students' total scale score, $r = .65$. The first task assessing students content knowledge, Task 2, had the strongest correlation with students' total scale score, $r = .87$. Overall, every one point gain in students' task scores corresponded to a 4.4 (Task 7) to 6.1 (Task 4) point gain in students' scale score.

***Grade 7.*** The regression of each content task total score on students' scale score was statistically significant for each task. Task 9 had the lowest correlation with students' total scale score, $r = .78$. Task 3 had the strongest correlation with students total scale score, $r = .89$. Overall, every one point gain in students' task scores corresponded to a 4.9 (Task 11) to 6.7 (Task 3) point gain in students' scale score.

***Grade 11.*** The regression of each content task total score on students' scale score was statistically significant for each task. Task 8 had the lowest correlation with students' total scale score, $r = .75$. Task 3 had the strongest correlation with students' total scale score, $r = .86$. Overall, every one point gain in students' task scores corresponded to a 5.2 (Task 2) to 6.6 (Task 4) point gain in students' scale score.

**Math**

***Grade 3.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 1054) = 2474.616$, MSR = 115016.111, $p < .05$, $R^2 = 0.70$. Task 2 total was a statistically significant predictor of students' scale score, $b = 2.780$, SE = .06, $p < .05$, 95% CI = 2.671 to 2.890. On average, every one point increase in the Task 2 total corresponded with a 2.780 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 1054) = 2474.616$, MSR = 115016.111, $p < .05$, $R^2 = 0.49$. Task 3 total was a statistically significant predictor of students' scale score, $b = 2.785$, SE = .09, $p < .05$, 95% CI = 2.614 to 2.955. On average, every one point increase in the Task 3 total corresponded with a 2.785 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 1054) = 2474.616$, MSR = 115016.111, $p < .05$, $R^2 = 0.49$. Task 4 total was a statistically significant predictor of students' scale score, $b = 2.785$, SE = .09, $p < .05$, 95% CI = 2.614 to 2.955. On average, every one point increase in the Task 4 total corresponded with a 2.785 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 1054) = 1118.559$, MSR = 84438.901, $p < .05$, $R^2 = 0.52$. Task 5 total was a statistically significant predictor of students' scale score, $b = 2.375$, SE = .07, $p < .05$, 95% CI = 2.236 to 2.515. On average, every one point increase in the Task 5 total corresponded with a 2.375 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 1054) = 1581.284$, MSR = 98409.667, $p < .05$, $R^2 = 0.60$. Task 6 total was a statistically significant predictor of students' scale score, $b = 2.535$, SE = .06, $p < .05$, 95% CI = 2.410 to 2.660. On average, every one point increase in the Task 6 total corresponded with a 2.535 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 1054) = 1937.057$, MSR = 106211.847, $p < .05$, $R^2 = 0.65$. Task 7 total was a statistically significant predictor of students' scale score, $b = 2.815$, SE = .06, $p < .05$, 95% CI = 2.689 to 2.940. On average, every one point increase in the Task 7 total corresponded with a 2.815 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 1054) = 1956.163$, MSR = 106578.680, $p < .05$, $R^2 = 0.65$. Task 8 total was a statistically significant predictor of students' scale score, $b = 2.609$, SE = .06, $p < .05$, 95% CI = 2.493 to 2.724. On average, every one point increase in the Task 8 total corresponded with a 2.609 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 1054) = 1764.283$, MSR = 102668.911, $p < .05$, $R^2 = 0.63$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.489$, SE = .06, $p < .05$, 95% CI = 2.373 to 2.606. On average, every one point increase in the Task 9 total corresponded with a 2.489 increase in students' scale scores.

***Grade 4.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 943) = 1876.673$, MSR = 131324.023, $p < .05$, $R^2 = 0.67$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.359$, SE = .08, $p < .05$, 95% CI =

3.207 to 3.511. On average, every one point increase in the Task 2 total corresponded with a 3.204 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 943) = 1134.661$, MSR = 107757.062, $p < .05$, $R^2 = 0.55$. Task 3 total was a statistically significant predictor of students' scale score, $b = 3.204$, SE = .10, $p < .05$, 95% CI = 3.018 to 3.391. On average, every one point increase in the Task 3 total corresponded with a 3.204 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 943) = 2034.870$, MSR = 134829.592, $p < .05$, $R^2 = 0.68$. Task 4 total was a statistically significant predictor of students' scale score, $b = 3.204$, SE = .07, $p < .05$, 95% CI = 3.136 to 3.421. On average, every one point increase in the Task 4 total corresponded with a 3.279 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 943) = 1701.729$, MSR = 126959.016, $p < .05$, $R^2 = 0.64$. Task 5 total was a statistically significant predictor of students' scale score, $b = 3.038$, SE = .07, $p < .05$, 95% CI = 2.894 to 3.183. On average, every one point increase in the Task 5 total corresponded with a 3.038 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 943) = 1680.486$, MSR = 126389.351, $p < .05$, $R^2 = 0.64$. Task 6 total was a statistically significant predictor of students' scale score, $b = 3.038$, SE = .07, $p < .05$, 95% CI = 2.884 to 3.174. On average, every one point increase in the Task 6 total corresponded with a 3.029 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 943) = 1995.364$, MSR = 133989.539, $p < .05$, $R^2 = 0.68$. Task 7 total was a statistically significant predictor of students' scale score, $b = 2.910$, SE = .07, $p < .05$, 95% CI = 2.670 to 2.952. On average, every one point increase in the Task 7 total corresponded with a 2.910 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 943) = 1538.280$, MSR = 122324.636, $p < .05$, $R^2 = 0.62$. Task 8 total was a statistically significant predictor of students' scale score, $b = 2.811$, SE = .07, $p < .05$, 95% CI = 2.782 to 2.811. On average, every one point increase in the Task 8 total corresponded with a 2.811 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 943) = 1298.581$, MSR = 114305.967, $p < .05$, $R^2 = 0.58$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.888$, SE = .08, $p < .05$, 95% CI = 2.730 to 3.045. On average, every one point increase in the Task 9 total corresponded with a 2.888 increase in students' scale scores.

***Grade 5.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 953) = 1155.557$, MSR = 96702.042, $p < .05$, $R^2 = 0.55$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.359$, SE = .08, $p < .05$, 95% CI = 3.207 to 3.511. On average, every one point increase in the Task 2 total corresponded with a 3.204 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 953) = 1134.661$, MSR = 107757.062, $p < .05$, $R^2 = 0.55$. Task 3 total was a statistically significant predictor of students' scale score, $b = 3.174$, SE = .09, $p < .05$, 95% CI = 2.990 to

3.357. On average, every one point increase in the Task 3 total corresponded with a 3.174 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 953) = 1306.911$, MSR = 102043.285, $p < .05$, $R^2 = 0.58$. Task 4 total was a statistically significant predictor of students' scale score, $b = 2.883$, SE = .09, $p < .05$, 95% CI = 2.726 to 3.039. On average, every one point increase in the Task 4 total corresponded with a 2.883 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 953) = 1201.438$, MSR = 98400.452, $p < .05$, $R^2 = 0.56$. Task 5 total was a statistically significant predictor of students' scale score, $b = 2.678$, SE = .08, $p < .05$, 95% CI = 2.526 to 2.829. On average, every one point increase in the Task 5 total corresponded with a 2.678 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 953) = 1427.877$, MSR = 105823.853, $p < .05$, $R^2 = 0.60$. Task 6 total was a statistically significant predictor of students' scale score, $b = 2.796$, SE = .08, $p < .05$, 95% CI = 2.650 to 2.941. On average, every one point increase in the Task 6 total corresponded with a 2.796 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 953) = 1240.864$, MSR = 99803.144, $p < .05$, $R^2 = 0.57$. Task 7 total was a statistically significant predictor of students' scale score, $b = 3.035$, SE = .09, $p < .05$, 95% CI = 2.866 to 3.205. On average, every one point increase in the Task 7 total corresponded with a 3.035 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 953) = 1528.197$, MSR = 108679.528, $p < .05$, $R^2 = 0.62$. Task 8 total was a statistically significant predictor of students' scale score, $b = 2.939$, SE = .08, $p < .05$, 95% CI = 2.791 to 3.086. On average, every one point increase in the Task 8 total corresponded with a 2.939 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 953) = 1851.362$, MSR = 116489.541, $p < .05$, $R^2 = 0.66$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.767$, SE = .06, $p < .05$, 95% CI = 2.640 to 2.893. On average, every one point increase in the Task 9 total corresponded with a 2.767 increase in students' scale scores.

***Grade 6.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 756) = 839.233$, MSR = 48087.430, $p < .05$, $R^2 = 0.53$. Task 2 total was a statistically significant predictor of students' scale score, $b = 2.435$, SE = .08, $p < .05$, 95% CI = 2.270 to 2.600. On average, every one point increase in the Task 2 total corresponded with a 2.435 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 756) = 1105.404$, MSR = 54281.715, $p < .05$, $R^2 = 0.59$. Task 3 total was a statistically significant predictor of students' scale score, $b = 2.274$, SE = .07, $p < .05$, 95% CI = 2.140 to 2.408. On average, every one point increase in the Task 3 total corresponded with a 2.274 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 756) = 503.499$, MSR = 36540.461, $p < .05$, $R^2 = 0.40$. Task 4 total was a statistically significant predictor of students' scale score, $b = 2.353$, SE = .11, $p < .05$, 95% CI = 2.147 to

2.559. On average, every one point increase in the Task 4 total corresponded with a 2.353 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 756) = 743.423$, MSR = 45319.501, $p < .05$, $R^2 = 0.50$. Task 5 total was a statistically significant predictor of students' scale score, $b = 2.279$, SE = .08, $p < .05$, 95% CI = 2.115 to 2.443. On average, every one point increase in the Task 5 total corresponded with a 2.279 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 756) = 1081.045$, MSR = 53789.459, $p < .05$, $R^2 = 0.59$. Task 6 total was a statistically significant predictor of students' scale score, $b = 2.488$, SE = .08, $p < .05$, 95% CI = 2.339 to 2.636. On average, every one point increase in the Task 6 total corresponded with a 2.488 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 756) = 886.904$, MSR = 49344.366, $p < .05$, $R^2 = 0.54$. Task 7 total was a statistically significant predictor of students' scale score, $b = 2.324$, SE = .08, $p < .05$, 95% CI = 2.171 to 2.477. On average, every one point increase in the Task 7 total corresponded with a 2.324 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 756) = 1014.650$, MSR = 52378.953, $p < .05$, $R^2 = 0.57$. Task 8 total was a statistically significant predictor of students' scale score, $b = 2.698$, SE = .09, $p < .05$, 95% CI = 2.532 to 2.864. On average, every one point increase in the Task 8 total corresponded with a 2.698 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 756) = 1262.888$, MSR = 57177.585, $p < .05$, $R^2 = 0.63$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.576$, SE = .07, $p < .05$, 95% CI = 2.433 to 2.718. On average, every one point increase in the Task 9 total corresponded with a 2.576 increase in students' scale scores.

***Grade 7.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 726) = 974.085$, MSR = 72722.771, $p < .05$, $R^2 = 0.57$. Task 2 total was a statistically significant predictor of students' scale score, $b = 2.789$, SE = .09, $p < .05$, 95% CI = 2.613 to 2.964. On average, every one point increase in the Task 2 total corresponded with a 2.789 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 726) = 1225.575$, MSR = 79707.451, $p < .05$, $R^2 = 0.63$. Task 3 total was a statistically significant predictor of students' scale score, $b = 2.964$, SE = .09, $p < .05$, 95% CI = 2.798 to 3.130. On average, every one point increase in the Task 3 total corresponded with a 2.964 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 726) = 1729.314$, MSR = 89394.552, $p < .05$, $R^2 = 0.70$. Task 4 total was a statistically significant predictor of students' scale score, $b = 2.742$, SE = .07, $p < .05$, 95% CI = 2.613 to 2.872. On average, every one point increase in the Task 4 total corresponded with a 2.742 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 726) = 1103.929$, MSR = 76568.692, $p < .05$, $R^2 = 0.60$. Task 5 total was a statistically significant predictor of students' scale score, $b = 2.933$, SE = .09, $p < .05$, 95% CI = 2.760 to

3.107. On average, every one point increase in the Task 5 total corresponded with a 2.933 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 726) = 1409.264$, MSR = 83769.336, $p < .05$, $R^2 = 0.66$. Task 6 total was a statistically significant predictor of students' scale score, $b = 2.623$, SE = .07, $p < .05$, 95% CI = 2.486 to 2.760. On average, every one point increase in the Task 6 total corresponded with a 2.623 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 726) = 597.231$, MSR = 57286.339, $p < .05$, $R^2 = 0.45$. Task 7 total was a statistically significant predictor of students' scale score, $b = 3.154$, SE = .13, $p < .05$, 95% CI = 2.901 to 3.408. On average, every one point increase in the Task 7 total corresponded with a 3.154 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 726) = 1114.753$, MSR = 76864.784, $p < .05$, $R^2 = 0.61$. Task 8 total was a statistically significant predictor of students' scale score, $b = 3.103$, SE = .09, $p < .05$, 95% CI = 2.920 to 3.285. On average, every one point increase in the Task 8 total corresponded with a 3.103 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 726) = 1095.963$, MSR = 76348.515, $p < .05$, $R^2 = 0.60$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.941$, SE = .09, $p < .05$, 95% CI = 2.767 to 3.116. On average, every one point increase in the Task 9 total corresponded with a 2.941 increase in students' scale scores.

***Grade 8.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 641) = 555.809$, MSR = 42507.147, $p < .05$, $R^2 = 0.46$. Task 2 total was a statistically significant predictor of students' scale score, $b = 2.612$, SE = .11, $p < .05$, 95% CI = 2.394 to 2.830. On average, every one point increase in the Task 2 total corresponded with a 2.612 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 641) = 644.364$, MSR = 45884.563, $p < .05$, $R^2 = 0.50$. Task 3 total was a statistically significant predictor of students' scale score, $b = 2.701$, SE = .11, $p < .05$, 95% CI = 2.492 to 2.909. On average, every one point increase in the Task 3 total corresponded with a 2.701 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 641) = 1072.154$, MSR = 57282.507, $p < .05$, $R^2 = 0.63$. Task 4 total was a statistically significant predictor of students' scale score, $b = 2.635$, SE = .08, $p < .05$, 95% CI = 2.477 to 2.793. On average, every one point increase in the Task 4 total corresponded with a 2.635 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 641) = 981.109$, MSR = 55360.317, $p < .05$, $R^2 = 0.61$. Task 5 total was a statistically significant predictor of students' scale score, $b = 2.432$, SE = .08, $p < .05$, 95% CI = 2.279 to 2.584. On average, every one point increase in the Task 5 total corresponded with a 2.432 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 641) = 1069.793$, MSR = 57235.258, $p < .05$, $R^2 = 0.69$. Task 6 total was a statistically significant predictor of students' scale score, $b = 2.560$, SE = .07, $p < .05$, 95% CI = 2.426 to

2.695. On average, every one point increase in the Task 6 total corresponded with a 2.560 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 641) = 1399.351$, MSR = 62774.471, $p < .05$, $R^2 = 0.63$. Task 7 total was a statistically significant predictor of students' scale score, $b = 2.963$, SE = .09, $p < .05$, 95% CI = 2.785 to 3.141. On average, every one point increase in the Task 7 total corresponded with a 2.963 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 641) = 993.853$, MSR = 55642.261, $p < .05$, $R^2 = 0.61$. Task 8 total was a statistically significant predictor of students' scale score, $b = 2.831$, SE = .09, $p < .05$, 95% CI = 2.655 to 3.007. On average, every one point increase in the Task 8 total corresponded with a 2.831 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 641) = 406.217$, MSR = 35504.458, $p < .05$, $R^2 = 0.39$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.856$, SE = .14, $p < .05$, 95% CI = 2.578 to 3.135. On average, every one point increase in the Task 9 total corresponded with a 2.856 increase in students' scale scores.

***Grade 11.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 354) = 421.331$, MSR = 53272.484, $p < .05$, $R^2 = 0.54$. Task 2 total was a statistically significant predictor of students' scale score, $b = 2$ 4.195, SE = .20, $p < .05$, 95% CI = 3.793 to 4.597. On average, every one point increase in the Task 2 total corresponded with a 4.195 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 354) = 306.797$, MSR = 45514.499, $p < .05$, $R^2 = 0.46$. Task 3 total was a statistically significant predictor of students' scale score, $b = 4.051$, SE = .23, $p < .05$, 95% CI = 3.596 to 4.506. On average, every one point increase in the Task 3 total corresponded with a 4.051 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 354) = 500.910$, MSR = 57438.880, $p < .05$, $R^2 = 0.59$. Task 4 total was a statistically significant predictor of students' scale score, $b = 3.943$, SE = .18, $p < .05$, 95% CI = 3.597 to 4.289. On average, every one point increase in the Task 4 total corresponded with a 3.943 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 354) = 336.159$, MSR = 47748.747, $p < .05$, $R^2 = 0.49$. Task 5 total was a statistically significant predictor of students' scale score, $b = 4.493$, SE = .25, $p < .05$, 95% CI = 4.011 to 4.975. On average, every one point increase in the Task 5 total corresponded with a 4.493 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 354) = 650.290$, MSR = 63476.714, $p < .05$, $R^2 = 0.65$. Task 6 total was a statistically significant predictor of students' scale score, $b = 3.481$, SE = .14, $p < .05$, 95% CI = 3.212 to 3.749. On average, every one point increase in the Task 6 total corresponded with a 3.481 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 354) = 334.307$, MSR = 47613.458, $p < .05$, $R^2 = 0.49$. Task 7 total was a statistically significant predictor of students' scale score, $b = 3.833$, SE = .21, $p < .05$, 95% CI = 3.421 to

4.245. On average, every one point increase in the Task 7 total corresponded with a 3.833 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 354) = 653.533$, MSR = 63587.953, $p < .05$, $R^2 = 0.65$. Task 8 total was a statistically significant predictor of students' scale score, $b = 3.355$, SE = .13, $p < .05$, 95% CI = 3.097 to 3.613. On average, every one point increase in the Task 8 total corresponded with a 3.355 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 354) = 621.430$, MSR = 62454.352, $p < .05$, $R^2 = 0.64$. Task 9 total was a statistically significant predictor of students' scale score, $b = 3.937$, SE = .16, $p < .05$, 95% CI = 3.627 to 4.248. On average, every one point increase in the Task 9 total corresponded with a 3.937 increase in students' scale scores.

**Science**

***Grade 5.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 565) = 1318.352$, MSR = 84096.722, $p < .05$, $R^2 = 0.70$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.870$, SE = .11, $p < .05$, 95% CI = 3.661 to 4.079. On average, every one point increase in the Task 2 total corresponded with a 3.870 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 565) = 1539.147$, MSR = 87878.615, $p < .05$, $R^2 = 0.73$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.786$, SE = .10, $p < .05$, 95% CI = 3.597 to 3.976. On average, every one point increase in the Task 3 total corresponded with a 3.786 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 565) = 1432.179$, MSR = 86150.838, $p < .05$, $R^2 = 0.72$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.585$, SE = .10, $p < .05$, 95% CI = 3.399 to 3.772. On average, every one point increase in the Task 4 total corresponded with a 3.585 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 565) = 1639.025$, MSR = 89340.477, $p < .05$, $R^2 = 0.74$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.585$, SE = .10, $p < .05$, 95% CI = 3.399 to 3.772. On average, every one point increase in the Task 5 total corresponded with a 3.585 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 565) = 1631.346$, MSR = 89232.808, $p < .05$, $R^2 = 0.74$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.575$, SE = .09, $p < .05$, 95% CI = 3.402 to 3.749. On average, every one point increase in the Task 6 total corresponded with a 3.575 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 565) = 1489.127$, MSR = 87093.080, $p < .05$, $R^2 = 0.73$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.612$, SE = .10, $p < .05$, 95% CI = 3.428 to 3.796. On average, every one point increase in the Task 7 total corresponded with a 3.612 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 565) = 1269.584$, MSR = 83138.657, $p < .05$, $R^2 = 0.69$. Task 2 total was a statistically

significant predictor of students' scale score, $b = 3.707$, SE = .10, $p < .05$, 95% CI = 3.503 to 3.911. On average, every one point increase in the Task 8 total corresponded with a 3.707 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 565) = 1501.226$, MSR = 87286.573, $p < .05$, $R^2 = 0.73$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.294$, SE = .09, $p < .05$, 95% CI = 3.127 to 3.461. On average, every one point increase in the Task 9 total corresponded with a 3.294 increase in students' scale scores.

***Grade 8.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 492) = 1009.573$, MSR = 59208.640, $p < .05$, $R^2 = 0.67$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.567$, SE = .11, $p < .05$, 95% CI = 3.347 to 3.788. On average, every one point increase in the Task 2 total corresponded with a 3.870 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 492) = 1297.174$, MSR = 63846.845, $p < .05$, $R^2 = 0.73$. Task 3 total was a statistically significant predictor of students' scale score, $b = 3.470$, SE = .10, $p < .05$, 95% CI = 3.281 to 3.659. On average, every one point increase in the Task 3 total corresponded with a 3.470 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 492) = 1139.943$, MSR = 61513.706, $p < .05$, $R^2 = 0.70$. Task 4 total was a statistically significant predictor of students' scale score, $b = 3.361$, SE = .10, $p < .05$, 95% CI = 3.165 to 3.557. On average, every one point increase in the Task 4 total corresponded with a 3.361 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 492) = 1018.283$, MSR = 59375.044, $p < .05$, $R^2 = 0.67$. Task 5 total was a statistically significant predictor of students' scale score, $b = 3.785$, SE = .12, $p < .05$, 95% CI = 3.552 to 4.018. On average, every one point increase in the Task 5 total corresponded with a 3.785 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 492) = 1012.473$, MSR = 59264.266, $p < .05$, $R^2 = 0.67$. Task 6 total was a statistically significant predictor of students' scale score, $b = 3.409$, SE = .11, $p < .05$, 95% CI = 3.198 to 3.619. On average, every one point increase in the Task 6 total corresponded with a 3.409 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 492) = 1073.608$, MSR = 60388.811, $p < .05$, $R^2 = 0.69$. Task 7 total was a statistically significant predictor of students' scale score, $b = 3.093$, SE = .09, $p < .05$, 95% CI = 2.907 to 3.278. On average, every one point increase in the Task 8 total corresponded with a 3.093 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 492) = 1161.333$, MSR = 61857.191, $p < .05$, $R^2 = 0.70$. Task 8 total was a statistically significant predictor of students' scale score, $b = 3.235$, SE = .10, $p < .05$, 95% CI = 3.049 to 3.422. On average, every one point increase in the Task 8 total corresponded with a 3.235 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 492) = 1245.419$, MSR = 63125.478, $p < .05$, $R^2 = 0.72$. Task 9 total was a statistically

significant predictor of students' scale score, $b = 3.020$, SE = .09, $p < .05$, 95% CI = 2.852 to 3.188. On average, every one point increase in the Task 9 total corresponded with a 3.020 increase in students' scale scores.

***Grade 11.*** The regression of **Task 2** total on students' scale score was statistically significant, $F(1, 279) = 519.617$, MSR = 84096.722, $p < .05$, $R^2 = 0.65$. Task 2 total was a statistically significant predictor of students' scale score, $b = 3.215$, SE = .14, $p < .05$, 95% CI = 2.937 to 3.493. On average, every one point increase in the Task 2 total corresponded with a 3.215 increase in students' scale scores.

The regression of **Task 3** total on students' scale score was statistically significant, $F(1, 279) = 581.039$, MSR = 29713.475, $p < .05$, $R^2 = 0.68$. Task 3 total was a statistically significant predictor of students' scale score, $b = 2.946$, SE = .12, $p < .05$, 95% CI = 2.705 to 3.186. On average, every one point increase in the Task 3 total corresponded with a 2.946 increase in students' scale scores.

The regression of **Task 4** total on students' scale score was statistically significant, $F(1, 279) = 388.483$, MSR = 25597.530, $p < .05$, $R^2 = 0.58$. Task 4 total was a statistically significant predictor of students' scale score, $b = 2.790$, SE = .14, $p < .05$, 95% CI = 2.512 to 3.069. On average, every one point increase in the Task 4 total corresponded with a 2.790 increase in students' scale scores.

The regression of **Task 5** total on students' scale score was statistically significant, $F(1, 279) = 599.120$, MSR = 30007.263, $p < .05$, $R^2 = 0.68$. Task 5 total was a statistically significant predictor of students' scale score, $b = 2.987$, SE = .12, $p < .05$, 95% CI = 2.747 to 3.228. On average, every one point increase in the Task 5 total corresponded with a 2.987 increase in students' scale scores.

The regression of **Task 6** total on students' scale score was statistically significant, $F(1, 279) = 435.456$, MSR = 26806.204, $p < .05$, $R^2 = 0.61$. Task 6 total was a statistically significant predictor of students' scale score, $b = 3.008$, SE = .14, $p < .05$, 95% CI = 2.725 to 3.292. On average, every one point increase in the Task 6 total corresponded with a 3.008 increase in students' scale scores.

The regression of **Task 7** total on students' scale score was statistically significant, $F(1, 279) = 546.504$, MSR = 29116.583, $p < .05$, $R^2 = 0.66$. Task 7 total was a statistically significant predictor of students' scale score, $b = 3.074$, SE = .13, $p < .05$, 95% CI = 2.816 to 3.333. On average, every one point increase in the Task 6 total corresponded with a 3.074 increase in students' scale scores.

The regression of **Task 8** total on students' scale score was statistically significant, $F(1, 279) = 703.463$, MSR = 31491.360, $p < .05$, $R^2 = 0.72$. Task 8 total was a statistically significant predictor of students' scale score, $b = 2.670$, SE = .10, $p < .05$, 95% CI = 2.472 to 2.869. On average, every one point increase in the Task 8 total corresponded with a 2.670 increase in students' scale scores.

The regression of **Task 9** total on students' scale score was statistically significant, $F(1, 279) = 428.286$, MSR = 26632.086, $p < .05$, $R^2 = 0.61$. Task 9 total was a statistically significant predictor of students' scale score, $b = 2.791$, SE = .14, $p < .05$, 95% CI = 2.526 to 3.057. On average, every one point increase in the Task 9 total corresponded with a 2.791 increase in students' scale scores.