

***WASHINGTON ALTERNATE ASSESSMENT SYSTEM
TECHNICAL REPORT
2007***

*Prepared for the
OFFICE OF THE SUPERINTENDENT OF PUBLIC INSTRUCTION*

by



October 2007

Table of Contents

I. EXECUTIVE SUMMARY----- 4

II. What is WAAS? ----- 5

 Program Purpose ----- 6

 Target Population & Participation Rates ----- 7

III. PORTFOLIO DEVELOPMENT ----- 8

IV. WAAS PORTFOLIO PROCESSES----- 10

 Pre-Range Finding and Range Finding ----- 10

 Summer Scoring Institute ----- 10

V. WAAS SCORING SCHEME AND PROCEDURES ----- 12

 General Scoring Scheme ----- 12

 Scoring Rules for Three Parts ----- 13

 Second Scoring and Resolution ----- 14

VI. TECHNICAL QUALITIES OF SCORING ----- 15

 Scoring of Qualifying Sets ----- 16

 Inter-Rater Agreement ----- 16

 Scoring of Validity Sets ----- 17

VII. CONSTRUCT VALIDITY----- 26

VIII. SUMMARY OF STUDENT PERFORMANCE ----- 33

 Analysis of Part I Scores ----- 34

 Analysis of Part II Scores ----- 34

 Total Score Distribution ----- 34

 Mean Score Summary ----- 35

IX. PERFORMANCE ON THE WAAS RELATIVE TO STANDARDS----- 43

APPENDIX A. WAAS PORTFOLIO SCORING 2007 SCORING HANDBOOK----- 48

APPENDIX B. WAAS PORTFOLIO SCORING SUMMARY SHEET—2007 ----- 64

APPENDIX C. WAAS PORTFOLIO ACADEMIC ACHIEVEMENT STANDARD DESCRIPTIONS ----- 65

List of Tables and Figures

Table 2.1 Number of Participants in the WAAS from 2002 to 2007	8
Table 5.1 Contents and Dimensions of the 2007 WAAS Scoring.....	13
Table 6.1 Results of the Scorer Qualification.....	18
Table 6.2 Percentage Agreement between First and Second Scorers for Grade 3	20
Table 6.3 Percentage Agreement between First and Second Scorers for Grade 4	20
Table 6.4 Percentage Agreement between First and Second Scorers for Grade 5	21
Table 6.5 Percentage Agreement between First and Second Scorers for Grade 6	21
Table 6.6 Percentage Agreement between First and Second Scorers for Grade 7	22
Table 6.7 Percentage Agreement between First and Second Scorers for Grade 8	22
Table 6.8 Percentage Agreement between First and Second Scorers for High School.....	23
Table 6.9 Overall Percentage Agreement between First and Second Scorers	23
Table 6.10 Scoring on the Validity Portfolios.....	24
Table 7.1 Grade 3 Sub-score Correlations	26
Table 7.2 Grade 4 Sub-score Correlations	27
Table 7.3 Grade 5 Sub-score Correlations	27
Table 7.4 Grade 6 Sub-score Correlations	28
Table 7.5 Grade 7 Sub-score Correlations	28
Table 7.6 Grade 8 Sub-score Correlations	29
Table 7.7 High School Sub-score Correlations	29
Table 7.8 Initial Principal Component Analysis for All Grades	31
Table 7.9 Factor Loadings in Fixed-Number Factor Analysis with Orthogonal Varimax Rotation...32	
Table 7.9 Factor Loadings in Fixed-Number Factor Analysis with Orthogonal Varimax Rotation (Continued).....	33
Table 8.1 2007 WAAS Score Distributions	35
Values in cells represent percentages.	35
Table 8.1 2007 WAAS Score Distributions (Continued)	36
Table 8.2 2007 WAAS Score Distribution Based on the Entire Population	37
Table 8.3 2007 WAAS Total Score Distributions.....	38
Table 8.3 2007 WAAS Total Score Distributions (Continued).....	39
Table 8.4 Mean Total Score Summary for All Grades and Contents	40
Table 8.4 Mean Total Score Summary for All Grades and Contents (Continued)	41
Table 8.4 Mean Total Score Summary for All Grades and Contents (Continued)	42
Table 9.1 Cut Scores and Percentages of Students in Each Category after Round 3.....	45
Table 9.2 Cut Scores and Percentages of Students in Each Category after Articulation	45
Table 9.3 Summary of Distribution by Performance Level on 2007 WAAS.....	46
Table 9.4 2007 WAAS Performance Distribution by Gender	47

I. EXECUTIVE SUMMARY

The Washington Alternate Assessment System (WAAS) was administered operationally for the seventh year in spring 2007. In compliance with professional standards that test developers produce a technical manual to document the technical quality of an assessment and evidence for the reliability and validity of test scores (*Standards for Educational and Psychological Testing*, AERA/APA/NCME, 1999), this manual summarizes the technical information for the 2007 WAAS.

The 2007 WAAS test administration window began the first day of the 2006 school year, and ended on March 30, 2007. All application materials were sent to the Pearson Scoring Center by this date. All received portfolios were scored during the Summer Scoring Institute convened from June 2 to June 12. Scored portfolios were returned to school districts in August. As WAAS continues to evolve and cover more grades and subject areas, the participation rate continues to grow. The total number of portfolios submitted increased from 4728 in 2006 to 5143 in 2007 (approximately 9% increase). Among the participants, 3300 (64%) were self-reported males and 1843 (36%) were self-reported females.

A modified scoring scheme was implemented beginning in 2007. Up to 2006, WAAS applied a two-part scoring system comprised of content skills (Part I) and generalization of content skills (Part II) for all submitted portfolios. In 2007, a three-part scoring method was introduced. Part I measures evidence of targeted skills. Four content areas in Part I are assessed in terms of the progress on IEP skills. Different from the prior years where only skill scores were assigned, alignment to Grade Level Expectation (GLE) and evidence of alignment to skills are now examined and evaluated in 2007. Part II is related to generalization of skills and concentrates mainly on evidence of setting and contextual evidence a student demonstrates in the portfolio, which was scored across all content areas in the previous years but is scored for each content area individually in 2007. In addition, a third dimension that concerns students' educational opportunities is examined across all content area. Part III includes three dimensions (modifications and adaptations, self determination, and interactions) that previously was a component of Part II scoring.

The technical quality of scores such as scorer reliability remains consistent from year to year. As in previous years, the inter-rater reliability is generally higher for Part I than for Part II. Although inter-correlations between Part I and Part II scores were fairly high and a dominating general factor is observed across all grades, statistical evidence also support the two-dimensional and content-specific construct of WAAS. Due to the new scoring rubrics and newly implemented standards, the 2007 score distributions are not directly comparable to prior years' distributions. A fairly large percentage of portfolios were identified as lacking sufficient evidence and they were given a score of zero, which affected the overall score average. However, for portfolios showing sufficient evidences of skills, the majority received the highest rating. The percentage of portfolios meeting standard remains consistent (between 55% and 75%) with the previous years.

II. What is WAAS?

State assessment programs provide one method of determining student academic achievement. The Washington State Assessment System provides accountability for program and educational opportunities for all students. Alternate assessment is one component of Washington's assessment system.

The Washington Alternate Assessment System (WAAS) program was developed by the Washington Alternate Assessment Task Force and expanded by Advisory Panels in response to requirements of the Individuals with Disabilities Education Act of 1997: "The State has established goals for the performance of children with disabilities in the state that . . . are consistent, to the maximum extent appropriate, with other goals and standards for children established by the state." The alternate assessments are based on Washington's Essential Academic Learning Requirements (EALRs) in the content areas of Communication, Reading, Writing, Mathematics, and Science. The state has prepared extensions for the EALRs. This document describes the critical function of the EALRs, the access skills, instructional activities, and assessment strategies that are designed to assist special education staff members to link functional IEP skills to the EALRs, to provide access to the general education curriculum, and to measure student progress toward achieving the EALRs. The most current version of the EALR extensions document can be found at http://www.k12.wa.us/CurriculumInstruct/EALR_GLE.aspx

A number of additional resources and documents can be downloaded from <http://www.k12.wa.us/SpecialEd/assessment.aspx>

The inclusion of students with disabilities in the assessment and accountability system is critical to ensure appropriate allocation of resources and learning opportunities for these students.

The Washington Alternate Assessment System was designed for a small percentage of the total school population. Students with disabilities are expected to take the Washington Assessment of Student Learning (WASL) tests, with or without necessary accommodations, unless the Individualized Education Program (IEP) team determines that the student is unable to participate on the WASL in one or more content areas. In this case, the IEP team may select the Washington Alternate Assessment System (WAAS) portfolio assessment.

Program Purpose

The Washington Alternate Assessment Task force—comprised of administrators, higher education personnel, teachers, and parents—determined the following two-fold purpose of the portfolio assessment:

- To provide an appropriate method of measuring progress on state goals and standards for students who are not able to access the WASL or any commercially available test, even with accommodations
- To ensure that students will be able to generalize the Individualized Education Program (IEP) skills to the maximum extent possible

The basic building block of the portfolio assessment is evidence of the student's performance and progress toward reaching IEP goals. Each of the entries in the portfolio documents two dimensions of learning: progress on IEP skills linked to the EALRs and student generalization of those skills.

Portfolio evidence should demonstrate participation in and progress toward IEP goals that are aligned to state standards (EALRs). In this way, evidence of progress on IEP skills is linked to the EALRs and can be linked to progress on state goals and standards.

Portfolio evidence should also show the extent to which a student can demonstrate and generalize the IEP skill linked to EALRs in the following ways:

- Use the IEP skill with appropriate modifications/adaptations, supports, or assistive technology in order to demonstrate all he or she knows and is able to do.
- Apply the IEP skill in a variety of settings and contexts in which the student is able to use learned skills. These places can include the classroom, other areas of the school, community settings, and home.
- Interact with non-disabled peers and others during IEP activities for the purpose of developing social relationships to enrich his or her life.
- Use self-determination skills in planning, monitoring, and evaluating IEP skill activities.

Target Population & Participation Rates

Federal guidelines indicate that states should develop alternate assessment participation guidelines so that approximately 2%–3% of the student population is eligible for an alternate assessment in each given year.

The increase in number of portfolios submitted in 2007 is mainly due to the increase in high school participation (from 532 in 2006 to 903 in 2007). In 2007, 5143 students participated in the WAAS program. As shown in Table 2.1, the participation rate of the WAAS remains consistent at almost all grade levels and shows a dramatic increase for high school participants. In 2002, only 427 portfolios across all grades were submitted. The number almost quadrupled in 2003 (1646), then nearly doubled to 3279 in 2005. The number reached a new high in 2007.

Table 2.1 Number of Participants in the WAAS from 2002 to 2007

Grade	Number of 2002 Participants	Number of 2003 Participants	Number of 2004 Participants	Number of 2005 Participants	Number of 2006 Participants	Number of 2007 Participants
3					720	774
4		695	726	780	801	785
5			389	605	685	724
6					630	616
7		425	531	678	680	625
8		174	440	589	680	716
HS		352	503	627	532	903
Total	427	1,646	2,589	3,279	4,728	5,143

III. PORTFOLIO DEVELOPMENT

The implementation of the WAAS portfolio is dependent on the interaction between the assessed student and the teacher or staff member who assists the student with portfolio construction. The teacher and the student must be cognizant of the components and types of evidence that are required and/or recommended for inclusion in the portfolio. The student must be able to demonstrate observable skills or to produce evidence to be included in the portfolio.

The teacher or staff member must be able to write measurable IEP goals or objectives that provide opportunities for the student to participate and progress in the general curriculum. Staff members must also be able to plan academic content-based activities and select one IEP skill linked

to EALRs that will be measured in each content area entry. Additionally, the assessment team that collects data (on the student's progress on IEP skills over time and the ability of the student to generalize and use these skills) must possess a certain level of assessment literacy about how best to measure assessment targets and document student growth in IEP skills.

Two ongoing activities have been implemented to document and control the effects of teacher knowledge of WAAS portfolio procedural issues or assessment practices. Regional teacher training sessions are conducted in the fall of each school year. Workshops are conducted in several regions of the state in the fall each year. The three-hour workshops cover implementing WAAS procedures, writing measurable IEP goals, planning general education content-based activities, and collecting student performance data. The WAAS portfolio session materials are posted to the OSPI Web site, and members of the Special Education Assessment Leadership Team (SEALT) are trained to replicate the WAAS workshops for teachers who missed an OSPI session. Participant surveys are conducted at the regional workshops to gauge perceptions of changes in instruction and assessment practices and to determine other training needs.

Research, data analyses, and reviews of the WAAS portfolios for a sample of copied portfolios each year has been the second strategy to inform us about the technical adequacy of the WAAS portfolio and to guide professional development for future WAAS workshops (Johnson & Arnold, 2004). Johnson addresses the following research questions in a 2004 report:

1. Does the WAAS include tasks from the extended benchmarks for the EALRs? If so, what kind and how frequently are they used?
2. How many low scores might be attributable to procedural issues with the portfolio?
3. In what way might we establish external validity of the WAAS in future administrations?

The findings of this study, perception surveys, and other studies were used to determine the research agenda for the 2007 WAAS portfolios review and to revise WAAS portfolio workshops to address professional learning needs.

IV. WAAS PORTFOLIO PROCESSES

Pre-Range Finding and Range Finding

Pre-range finding occurred from May 15 to May 17, 2007. All grades (3-12) and their corresponding contents were reviewed and assessed. Fourteen districts had been identified for pre-range finding, with approximately 300 portfolios to select from. The objective was to select 30 to 40 portfolios to be reviewed at subsequent range-finding.

The portfolio range finding occurred over a one-week period in May (May 21–25) following the pre-range finding. During the week, a group of representatives from ILSSA, Pearson, and Washington educators were led by OSPI staff to review the anchor portfolios and prepare for scoring. ILSSA and Pearson staff have many years of experience scoring the alternate assessment. Using the 30 to 40 portfolios selected during pre-range finding, approximately 10 portfolios were chosen for training. Among them one was used as anchor set, one as practice set, two as qualifiers, and three as validity portfolios. The anchor portfolios, exemplified score points for the rubric. Scoring summaries and annotations were written to accompany the training sets for the Summer Scoring Institute.

Summer Scoring Institute

Portfolio scoring activities started the week after range finding. On May 30 and 31, table leader training for alignment scoring was conducted. The table leaders are reliable and experienced scorers from previous years. During this training, table leaders were trained to score for alignment with an anchor set and practice set. Table leaders had to qualify on an alignment set to be a table leader. Alignment scoring occurred at two different times. The first time was immediately after training and the second time was after the first group of teacher scorers completed their scoring session.

Summer Scoring Institute immediately followed the first session of alignment scoring. The first session of teacher scoring began on June 2 and lasted until June 5. The second session of teacher scoring began on June 9 and lasted until June 12. Interested teachers completed an OSPI application to become WAAS portfolio scorers.

During Summer Scoring Institute, OSPI assigned scorers to each table, and table leaders were randomly assigned to the tables. ILSSA leadership staff served as scoring trainers. Most educators participated during both weeks of scoring, but some scorers during the Summer Scoring Institute were new; therefore, the training process was repeated for the second session. The first day of each session was dedicated to a full day of training. During the second session, returning teachers did not participate in the training but continued scoring after taking a validity portfolio.

Participating scorers were trained to use the rubric using the exemplar portfolios. They were trained to score independently using two exemplar portfolios (an anchor portfolio and a practice portfolio). When OSPI, ILLSA and Pearson concluded that all teachers were properly trained, scoring procedures were reviewed. After training, the teachers scored a qualifying portfolio. Teachers who did not qualify after the first qualifying portfolio were given an opportunity to score a second qualifying portfolio.

Appendix A describes the scoring procedures. All scorers completed first and second scoring. The scorer logged the portfolio on the Batch Tracking Log. Scorers were not allowed to select a portfolio from their own school district or from school districts of their tablemates to score. Scorers used the Scoring Summary Sheets (Appendix B) to record scores and then fill in the scoring monitors. After scoring the portfolio, the scorers placed the monitor in a tray on the table leader's desk and put the completed portfolio in a box labeled "To Be Filed" (if the portfolio had only a first score monitor) or "To Be Second Scored" (if the portfolio had a second score monitor). If the portfolio required a second score, it was then placed at a different table to be second scored, so that the same team of scorers did not first and second score the same portfolio.

Table leaders and scorers completed a validity portfolio on the second day of scoring and after a break from scoring. The leadership team also reviewed all reliability statistics, including validity agreement, each afternoon.

V. WAAS SCORING SCHEME AND PROCEDURES

General Scoring Scheme

A new scoring construct was implemented beginning in 2007. Different from previous years where portfolios were scored in only two parts, a three-part scoring method was applied in 2007.

As in previous years, Part I examines progress on IEP skills and its scores are determined by evidence in separate portfolio entries for Reading, Writing, Mathematics, and Science. Instead of having only one skill score, there may be two skills defined in a specific content area. The single Part I content skill score is now extended to include 1) alignment to a GLE that is scored dichotomously as “Yes” or “No,” 2) evidence of alignment to specific skill that is on a five-point scale (1=very little evidence aligned to the target skill, 2=some evidence, 3=most evidence, 4=all evidence aligned, and insufficient evidence, or IE), and 3) skill score that is on the same five-point scale as evidence of alignment (1= not approaching the goal on target skill, 2=approached the goal, 3=met the goal, 4=exceeded the goal, or IE).

Part II scoring examines generalization of content skills. This part used to contain four scoring dimensions including Setting and Contexts, Modifications and Adaptations, Social Relations, and Self-Determination that were determined by examining evidence across all content areas in a portfolio. In 2007, Part II scoring was reduced to “Setting and Contexts.” This dimension is evaluated and scored for each individual content area on a three-point scale (1, 2, and insufficient evidence).

The remaining three dimensions are grouped in Part III scoring that examines student’s educational opportunities. While definitions of “Modifications and Adaptations” and “Self-Determination” remain unchanged, “Social Relations” is renamed as “Interaction” and redefined to examine evidence of student’s varied interactions with others related to the academic skills and/or in work, social, and/or community situations. Each of the three Part III dimensions is scored across all content areas on a three-point scale (1, 2, and insufficient evidence).

The scoring scheme is designed to only look for scores for appropriate content areas depending on the student’s grade. No score will be presented for content areas not required at a grade level. This produces two total scores for grades 3 and 6 (Reading and Mathematics), three

separate total scores (Reading, Writing, and Mathematics) for the students in grades 4 and 7, three total scores (Reading, Mathematics, and Science) for grades 5 and 8, and four scores for high school (Reading, Writing, Mathematics, and Science). Table 5.1 lists the various scoring dimensions and the applicable grades. Appendix B includes the scoring sheet used during the process. The scoring sheet also provides definitions for each score category.

Table 5.1 Contents and Dimensions of the 2007 WAAS Scoring

	IEP Skills	Grades Assessed
Part I	Mathematics	All
	Reading	All
	Science	5, 8, HS
	Writing	4, 7, HS
Part II	Settings and Contexts	All
Part III	Modifications and Adaptations	All
	Self Determination	All
	Interactions	All

Scoring Rules for Three Parts

As mentioned, Part I may have up to two defined skills. If only one skill is present, the total skill score is calculated by adding the content skill (Part I) and the setting (Part II) scores and doubled, whereas if two skills are present in a content area, the higher of the two total skill scores (Part I + Part II) is doubled. This rule applies to all grades. Other general scoring rules are described below in detail:

- If only one skill is submitted for a required content area and the skill score for the content area is deemed an IE, Settings and Contexts must also receive a score of IE for that content area. If two skills are submitted and one skill score is IE and the other skill score is numeric,

the Settings and Contexts score may be numeric or IE. If both skill scores within a content area are IE, the Settings and Contexts score must be IE.

- If at least one skill score within a content area receives a numeric score, Part III (opportunity score) may receive a numeric score or an IE. However, if all required skills scores receive scores of IE or nothing, the opportunity scores must also be IE.
- The scoring system will systematically convert all IE score to “0.” This will allow the scoring system to sum the Part I skill score and Part II setting score within each content area. However, IE will still be printed on paper reports.
- If all required content areas have evidence and skill scores of PP (Previously Passed), RF (Refusal), or NT (Content Area Not Tested), the Part III opportunity scores should also reflect the condition.
- If any one of the required content areas receives a score of IE and all other content areas receive any combination of other condition codes (PP, RF, NT), the Part III opportunity scores should be IE.

After scoring all three parts, a total score for each content area is generated by doubling the sum of Part I skill score and Part II Settings and Contexts score.

Second Scoring and Resolution

The first and second scoring monitors were compared through the ePS system. For grades 3–8, approximately 25% of the portfolios are second scored, but no resolution is required. In all cases, regardless of adjacency, the first score is used as the student’s final score.

For high school (grade 9-12), all portfolios are second scored. The system checks each score for adjacency. In the case of a non-adjacent score, only that score is resolved. In other words, resolution occurs at the line level for only the skill scores and setting scores. For example, if reading skill scores are adjacent and reading setting scores are not, only the setting scores require resolution. Alignment and evidence scores are not reviewed for adjacency. If the two scores are adjacent and no resolution is required, the first score overrides the second score. If a third score (resolution score) is required, the resolution score replaces all other scores for that score line and is used for reporting.

Table leaders back read each scorer every day. Those scorers with lower reliability or validity were back read more often. Table leaders use back reading as a method for monitoring scorer's accuracy. The results of back reading were recorded on back read summary sheets. Inter-rater reliability statistics are calculated each day. Whole-group recalibration and training occurs for any scoring dimension statistic that deviates from other dimensions or previous years. Individual scorer retraining occurs for those scorers with less than 66% exact or adjacent scores on validity portfolios.

Table leaders and scorers are asked to score a validity portfolio, which is a portfolio that has been previously scored by OSPI/Pearson/ILLSA leadership staff. The leadership staff review the results for scorers or table leaders who have less than 66% exact matches and adjacent scores. Scorers with lower percentages of inter-rater agreement are retrained.

To ensure that students received accurate scoring judgments, these two procedures are followed:

- Table leaders “backread” at least one portfolio from each scorer at their table daily.
- Reliability of scorers is monitored by comparing scorer performance relative to second reads and by reviewing their performance on validity portfolios.

The WAAS Portfolio Scoring Summary is attached in Appendix B. It illustrates the scoring scheme described above and the scoring rubrics scorers used to determine the portfolio quality.

VI. TECHNICAL QUALITIES OF SCORING

In this section, qualities of the 2007 WAAS portfolio scores are examined through the analysis of variance between and within scorers. Three methods are used to examine the quality of WAAS scoring. First is the scorers' performance on the qualifying sets. Second is the inter-rater agreement between scorers. The third method to evaluate scoring quality is scoring of the pre-selected validity papers.

Scoring of Qualifying Sets

Before actually scoring the portfolios, scorers (both table leaders and regular scorers) independently score either one or two qualification sets to determine that they fully grasp the scoring criteria and that they can assign scores correctly and consistently according to the specifications of WAAS. There were two pre-selected qualification sets. Each qualification set contained one complete portfolio. Scorers were required to achieve a certain score on one of the two qualifying sets in order to qualify to begin scoring. The percentage required to score was established by OSPI. Scorers not meeting the established guidelines by the end of the training sessions (or after scoring the two qualifying sets) were placed on probation or dismissed. As shown in Table 6.1 below, only three scorers had to take a second qualifying portfolio, and all scorers achieved the required level of accuracy after the second qualifying set.

Inter-Rater Agreement

Inter-rater agreement is an important source of evidence for the reliability of test scores. When two trained judges agree with the score given to a student's work, it supports the concept that this is the "correct" score for that student's work. The percent of agreement between scorers is examined to determine the degree to which judges give equivalent scores to the same student work. Inter-rater agreement of the first and second scorers is monitored on daily basis. Tables 6.2–6.9 summarize the extent of agreement between the first and second scorers for each scoring dimension at each grade over five days of scoring. As shown in Table 6.9, the overall percentage of exact agreement or of adjacent scores for all four content areas (Part I) is fairly high, ranging from 92.5% to 93.5%. This is also true when the analysis is performed independently for each grade (Tables 6.2–6.8). On the other hand, Part II scores generally exhibit lower inter-rater agreement relative to Part I scores. Part III scores show fairly high level of inter-rater agreement. They are generally higher than the inter-rater agreement of Part II scores, especially for higher grade levels. While Part II scoring seems to be less reliable than Part I or Part III, the majority of the Part II scores exhibit exact agreement regardless of grade level and content area. Referring again to the overall result in Table 6.9, although rates of perfect agreement for Part II scores are not as high for Part I or Part III, they still range from 72% to 77%, which may be considered satisfactory. The difference in score agreement may be explained by different interpretations by raters of the Part II scoring definition

and rubrics and limited evidence for Part II scores. These results are, however, consistent with those from 2002 and 2006. While there is variation in agreement across scoring dimensions, these percentages of agreement appear to be fairly reasonable.

Scoring of Validity Sets

The result from scoring validity portfolios was also monitored. There were three total portfolios in the validity set (V1, V2, and V3). Through the scoring process, the validity portfolios were administered to all scorers and table leaders on certain days (two administered on the second day of scoring for each group of scorers, and one administered to the returning group of scorers after a three-day lay-off). Scorers and tables leaders were assigned up to all three validity portfolios. A total of 141 validity portfolio reads were collected. Among them 31 were from table leaders and 110 were from regular scorers. All validity portfolios were from grade 10 because grade 10 requires scores for all content areas.

This process was used as a quality check to ensure that scorers had not drifted in their understanding of the scoring criteria and were continuing to score accurately. Validity portfolios, like qualification portfolios, were chosen from those portfolios scored by the range-finding committee and were strong exemplars endorsed by OSPI. Validity materials were developed by Pearson and ILSSA and approved by OSPI. All validity portfolios were distributed at the same time to all scorers and table leaders (at the beginning of the day) and were scored independently.

Results of the scoring on the validity portfolios are presented in Table 6.10.

Table 6.1 Results of the Scorer Qualification

Table Leaders			Regular Scorers		
ID	Average Agreement on the First Qualifying Portfolio	Average Agreement on the Second Qualifying Portfolio	ID	Average Agreement on the First Qualifying Portfolio	Average Agreement on the Second Qualifying Portfolio
1009	83%		2021	91%	
1010	50%	83%	2009	33%	100%
1003	100%		3017	25%	100%
1011	83%		3010	66%	
1004	91%		3014	91%	
1007	100%		3028	91%	
1008	66%		3013	83%	
1001	91%		2022	66%	
1002	83%		3029	100%	
1005	66%		3001	83%	
1006	66%		2008	83%	
			2025	100%	
			3026	100%	
			3018	83%	
			3019	75%	
			3016	83%	
			2016	91%	
			2001	100%	
			2000	91%	
			3008	66%	
			2003	75%	
			3015	83%	
			3005	83%	
			2002	83%	
			3003	83%	
			3000	83%	
			3030	100%	
			3021	75%	
			3031	66%	
			2005	100%	
			2020	66%	
			2007	83%	
			3024	83%	
			2023	91%	
			2015	75%	
			2012	100%	

Table Leaders			Regular Scorers		
Scorer	Average Agreement on the First Qualifying Portfolio	Average Agreement on the Second Qualifying Portfolio	Scorer	Average Agreement on the First Qualifying Portfolio	Average Agreement on the Second Qualifying Portfolio
			4002	83%	
			4007	91%	
			4008	91%	
			4005	83%	
			4012	83%	
			4009	83%	
			4001	100%	
			4004	91%	
			4003	100%	
			4010	83%	
			2019	83%	
			3020	75%	
			3011	66%	
			2018	100%	
			2017	91%	
			2024	83%	
			3002	100%	
			2014	83%	
			3004	91%	
			3027	91%	
			3007	75%	
			3023	75%	
			3006	100%	
			2004	100%	
			3022	75%	
			3009	91%	
			2006	91%	

Table 6.4 Percentage Agreement between First and Second Scorers for Grade 5

Extent of Agreement	Part I: Content Skills				Part II: Generalization of the Content Skills (Setting and Contexts)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Perfect Agreement	87.22%		85.71%	85.85%	68.04%		65.12%	70.37%	89.66%	88.35%	89.62%
Adjacent	8.27%		11.11%	11.32%	31.96%		34.88%	29.63%	10.34%	11.65%	10.38%
Non-Adjacent	4.51%		3.17%	2.83%							

Table 6.5 Percentage Agreement between First and Second Scorers for Grade 6

Extent of Agreement	Part I: Content Skills				Part II: Generalization of the Content Skills (Setting and Contexts)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Perfect Agreement	86.96%		80.53%		86.36%		84.52%		95.18%	92.94%	93.68%
Adjacent	6.96%		12.39%		13.64%		15.48%		4.82%	7.06%	6.32%
Non-Adjacent	6.09%		7.08%								

Table 6.6 Percentage Agreement between First and Second Scorers for Grade 7

Extent of Agreement	Part I: Content Skills				Part II: Generalization of the Content Skills (Setting and Contexts)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Perfect Agreement	79.35%	78.89%	88.30%		80.65%	81.36%	90.77%		85.51%	93.42%	90.12%
Adjacent	14.13%	15.56%	10.64%		19.35%	18.64%	9.23%		14.49%	6.58%	9.88%
Non-Adjacent	6.52%	5.55%	1.06%								

Table 6.7 Percentage Agreement between First and Second Scorers for Grade 8

Extent of Agreement	Part I: Content Skills				Part II: Generalization of the Content Skills (Setting and Contexts)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Perfect Agreement	81.42%		80.00%	75.73%	81.25%		70.51%	74.65%	89.16%	95.65%	90.91%
Adjacent	12.39%		13.64%	16.50%	18.75%		29.49%	25.35%	10.84%	4.35%	9.09%
Non-Adjacent	6.19%		6.36%	7.76%							

Table 6.8 Percentage Agreement between First and Second Scorers for High School

Extent of Agreement	Part I: Content Skills				Part II: Generalization of the Content Skills (Setting and Contexts)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Perfect Agreement	79.74%	77.74%	82.74%	80.19%	77.22%	71.94%	75.84%	74.24%	96.18%	97.96%	96.28%
Adjacent	13.34%	13.67%	9.84%	12.12%	22.78%	28.06%	24.16%	25.76%	3.82%	2.04%	3.72%
Non-Adjacent	6.91%	8.58%	7.42%	7.69%							

Table 6.9 Overall Percentage Agreement between First and Second Scorers

Extent of Agreement	Part I: Content Skills				Part II: Generalization of the Content Skills (Setting and Contexts)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Perfect Agreement	81.41%	77.88%	83.27%	80.38%	77.01%	71.85%	75.51%	73.65%	90.89%	94.65%	91.97%
Adjacent	11.60%	14.63%	10.21%	12.62%	22.99%	28.15%	24.49%	26.35%	9.11%	5.35%	8.03%
Non-Adjacent	6.99%	7.49%	6.52%	6.99%							

Table 6.10 Scoring on the Validity Portfolios

Table Leaders					Regular Scorers				
ID	Average Agreement on V1	Average Agreement on V2	Average Agreement on V3	Cumulative Agreement on All Validity Portfolios	ID	Average Agreement on V1	Average Agreement on V2	Average Agreement on V3	Cumulative Agreement on All Validity Portfolios
1009	75%	75%	75%	75%	2021	58%	91%	91%	80%
1010	83%	91%	91%	88%	2009	75%	75%	91%	80%
1003	91%	91%	100%	94%	3017	83%			83%
1011	100%	100%	100%	100%	3010	66%			66%
1004	75%			75%	3014	91%			91%
1007	91%	75%	100%	88%	3028	100%			100%
1008	83%	91%	91%	88%	3013	83%			83%
1001	91%	75%	91%	85%	2022	83%	91%	91%	88%
1002	91%	83%	75%	83%	3029	91%			91%
1005	83%	66%	100%	83%	3001	66%			66%
1006	83%	75%	100%	86%	2008	83%	100%	100%	94%
					2025	83%	83%	83%	83%
					3026	83%			83%
					3018	83%			83%
					3019	83%			83%
					3016	83%	75%	91%	83%
					2016	58%	66%	91%	71%
					2001	83%	91%	91%	88%
					2000	91%	91%	91%	91%
					3008	75%			75%
					2003	83%	91%	91%	88%
					2019	75%	75%	75%	75%
					3020	83%			83%
					3011	75%			75%
					2018	100%	83%	91%	91%
					2017	91%	91%	91%	91%
					2024	75%	83%	50%	69%
					3002	58%			58%
					2014	83%	91%	100%	91%
					3004	83%			83%
					3027	83%			83%
					3007	83%			83%
					3023	58%			58%
					3006	100%			100%
					2004	83%	91%	91%	88%
					3022	75%			75%
					3009	91%			91%
					2006	75%		100%	87%
					3015	100%			100%

Table Leaders					Regular Scorers				
Scorer	Average Agreement on V1	Average Agreement on V2	Average Agreement on V3	Cumulative Agreement on All Validity Portfolios	Scorer	Average Agreement on V1	Average Agreement on V2	Average Agreement on V3	Cumulative Agreement on All Validity Portfolios
					3005	66%			66%
					2002	83%	83%	83%	83%
					3003	50%			50%
					3000	66%			66%
					3030	83%			83%
					3021	83%			83%
					3031	75%			75%
					2005	83%	91%	58%	77%
					2020	75%	91%	58%	74%
					2007	83%	100%	83%	88%
					3024	83%			83%
					2023	75%	66%	91%	77%
					2015	91%	75%	100%	88%
					2012	100%	83%	91%	91%
					4002			75%	75%
					4007			83%	83%
					4008			91%	91%
					4005			91%	91%
					4012			91%	91%
					4009			100%	100%
					4001			100%	100%
					4004			91%	91%
					4003			83%	83%
					4010			91%	91%

VII. CONSTRUCT VALIDITY

Common approaches to investigate construct validity of an assessment are to examine inter-dimension correlations and factor structure. The correlation method provides information about the relationships among the test dimensions. Similarly, the factorial method explores the structure of an assessment through the correlations among the sub-scores.

Tables 7.1 to 7.7 show the correlations among the sub-scores for all the grades. The correlations within Part I, Part II, and Part III are generally higher than the correlations between the three parts, except when correlating within the same content domain. For example, Correlations among reading, writing, math, and science skill scores (Part I) are usually higher than correlations between reading skill score (Part I) and writing, math, and science setting scores (Part II). However, correlation between reading skill score (Part I) and reading setting score (Part II) is high. Such inter-sub-score correlations can be observed across all grades and support the three-part scoring structure of the WAAS. The correlation matrices suggest that generalization of the content skill (contexts and setting in Part II scoring) and students' educational opportunities (Part III scoring) are different yet reasonably relevant measures to the content skill (Part I) itself.

Table 7.1 Grade 3 Sub-score Correlations

Dimension	Part I		Part II		Part III		
	Reading	Math	Reading Setting	Math Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00						
Math	0.52	1.00					
Reading Setting	0.68	0.45	1.00				
Math Setting	0.42	0.69	0.59	1.00			
Modifications and Adaptations	0.51	0.51	0.56	0.57	1.00		
Self Determination	0.55	0.55	0.58	0.56	0.64	1.00	
Interactions	0.57	0.58	0.66	0.61	0.67	0.69	1.00

Table 7.2 Grade 4 Sub-score Correlations

Dimension	Part I			Part II			Part III		
	Reading	Writing	Math	Reading Setting	Writing Setting	Math Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00								
Writing	0.56	1.00							
Math	0.48	0.56	1.00						
Reading Setting	0.68	0.45	0.36	1.00					
Writing Setting	0.42	0.67	0.42	0.53	1.00				
Math Setting	0.41	0.48	0.66	0.52	0.59	1.00			
Modifications and Adaptations	0.47	0.42	0.37	0.56	0.51	0.46	1.00		
Self Determination	0.58	0.51	0.43	0.60	0.51	0.48	0.62	1.00	
Interactions	0.54	0.52	0.44	0.61	0.59	0.53	0.65	0.72	1.00

Table 7.3 Grade 5 Sub-score Correlations

Dimension	Part I			Part II			Part III		
	Reading	Math	Science	Reading Setting	Math Setting	Science Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00								
Math	0.55	1.00							
Science	0.53	0.56	1.00						
Reading Setting	0.66	0.47	0.44	1.00					
Math Setting	0.43	0.65	0.42	0.60	1.00				
Science Setting	0.43	0.48	0.72	0.56	0.58	1.00			
Modifications and Adaptations	0.48	0.43	0.46	0.52	0.46	0.52	1.00		
Self Determination	0.55	0.50	0.56	0.55	0.50	0.57	0.70	1.00	
Interactions	0.57	0.53	0.53	0.61	0.58	0.60	0.72	0.76	1.00

Table 7.4 Grade 6 Sub-score Correlations

Dimension	Part I		Part II		Part III		
	Reading	Math	Reading Setting	Math Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00						
Math	0.65	1.00					
Reading Setting	0.65	0.50	1.00				
Math Setting	0.51	0.63	0.63	1.00			
Modifications and Adaptations	0.51	0.47	0.53	0.55	1.00		
Self Determination	0.58	0.55	0.55	0.54	0.62	1.00	
Interactions	0.61	0.59	0.63	0.67	0.67	0.72	1.00

Table 7.5 Grade 7 Sub-score Correlations

Dimension	Part I			Part II			Part III		
	Reading	Writing	Math	Reading Setting	Writing Setting	Math Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00								
Writing	0.61	1.00							
Math	0.67	0.64	1.00						
Reading Setting	0.71	0.49	0.53	1.00					
Writing Setting	0.51	0.65	0.47	0.64	1.00				
Math Setting	0.57	0.51	0.72	0.60	0.60	1.00			
Modifications and Adaptations	0.52	0.54	0.56	0.56	0.54	0.57	1.00		
Self Determination	0.58	0.61	0.65	0.55	0.55	0.56	0.62	1.00	
Interactions	0.60	0.57	0.60	0.63	0.62	0.63	0.69	0.75	1.00

Table 7.6 Grade 8 Sub-score Correlations

Dimension	Part I			Part II			Part III		
	Reading	Math	Science	Reading Setting	Math Setting	Science Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00								
Math	0.69	1.00							
Science	0.47	0.41	1.00						
Reading Setting	0.68	0.53	0.40	1.00					
Math Setting	0.51	0.66	0.35	0.61	1.00				
Science Setting	0.36	0.36	0.67	0.52	0.49	1.00			
Modifications and Adaptations	0.39	0.39	0.47	0.44	0.47	0.51	1.00		
Self Determination	0.49	0.46	0.56	0.53	0.51	0.58	0.67	1.00	
Interactions	0.48	0.49	0.53	0.53	0.56	0.59	0.62	0.76	1.00

Table 7.7 High School Sub-score Correlations

Dimension	Part I				Part II				Part III		
	Reading	Writing	Math	Science	Reading Setting	Writing Setting	Math Setting	Science Setting	Modifications and Adaptations	Self Determination	Interactions
Reading	1.00										
Writing	0.57	1.00									
Math	0.61	0.60	1.00								
Science	0.59	0.58	0.65	1.00							
Reading Setting	0.62	0.35	0.44	0.44	1.00						
Writing Setting	0.42	0.59	0.40	0.42	0.55	1.00					
Math Setting	0.42	0.37	0.63	0.46	0.55	0.54	1.00				
Science Setting	0.44	0.40	0.48	0.67	0.54	0.52	0.55	1.00			
Modifications and Adaptations	0.39	0.40	0.40	0.37	0.43	0.41	0.44	0.44	1.00		
Self Determination	0.46	0.46	0.50	0.50	0.43	0.45	0.45	0.44	0.61	1.00	
Interactions	0.44	0.42	0.47	0.43	0.48	0.48	0.52	0.47	0.61	0.68	1.00

Exploratory factor analyses were performed to further investigate the nature of the inter-correlations across the score dimensions. A principle components analysis was conducted using SAS (Statistical Analysis Software, v 9.1). The number of factors was determined using three criteria: eigenvalues greater than 1, a scree test for the eigenvalues, and finding the solution in which approximately 40% of the variance was explained. Table 7.8 lists the eigenvalues and proportion of variance explained of all the extracted components. The result was a one-factor solution for all grades, where 54%–65% percent of the variance was explained. This result is consistent with the analysis of the inter-correlations.

While the principal component analysis suggests a single-factor solution and that 2007 WAAS scores are well correlated within each grade, it is also of interest to examine the multi-factorial structure on the basis of content domains. Based on the a priori hypothesis that there should be two to four distinct factors (depending on the content domains in a specific grade), a multi-factor solution was also investigated for each individual grade by fixing the number of factor to 4. An Orthogonal Varimax rotation approach was applied in the analysis.

Patterns of factor loadings in Table 7.9 clearly suggest dimension- and domain-specific factorial structure underlying the general factor. In most grades (grade 3 to 8 namely), we observe clustered loadings of Part III scores on Factor 1. This indicates Part III dimensions are related yet distinct from the content skills and generalization of the skills. Following the first factor, one clearly sees clustered loadings on each individual content domain. In grade 3, high loadings on the second factor are math-related scores, and reading-related scores load on the third factor. Similar patterns are observed in every grade level through grade 8.

The factor pattern for high school shows a strong dimension-specific structure. The first factor seems to represent the content skills as loading from Part I are high. Part III opportunity scores load high on actor 2, whereas Part II settings scores load high on the third factor.

Regardless of grade levels, factor loadings of the multi-factor solution show that scores of different dimensions (parts) and content domains have distinct loading patterns from each other, reflecting the findings from the correlation analysis. While scores of WAAS are correlated to a certain degree, they seem to provide distinct and unique information about students' performance.

Table 7.8 Initial Principal Component Analysis for All Grades

Component	Eigenvalue	Proportion	Component	Eigenvalue	Proportion
Grade 3			Grade 7		
1	4.48	0.64	1	5.77	0.64
2	0.70	0.10	2	0.61	0.07
3	0.55	0.08	3	0.61	0.07
4	0.43	0.06	4	0.56	0.06
5	0.36	0.05	5	0.49	0.05
6	0.29	0.04	6	0.38	0.04
7	0.19	0.03	7	0.22	0.02
Grade 4			8	0.19	0.02
1	5.21	0.58	9	0.17	0.02
2	0.93	0.10	Grade 8		
3	0.67	0.07	1	5.17	0.57
4	0.62	0.07	2	1.08	0.12
5	0.50	0.06	3	0.68	0.08
6	0.38	0.04	4	0.54	0.06
7	0.27	0.03	5	0.50	0.06
8	0.22	0.02	6	0.38	0.04
9	0.20	0.02	7	0.24	0.03
Grade 5			8	0.22	0.02
1	5.40	0.60	9	0.18	0.02
2	0.79	0.09	High School		
3	0.72	0.08	1	5.91	0.54
4	0.63	0.07	2	1.01	0.09
5	0.53	0.06	3	0.83	0.08
6	0.30	0.03	4	0.69	0.06
7	0.23	0.03	5	0.64	0.06
8	0.22	0.02	6	0.58	0.05
9	0.18	0.02	7	0.41	0.04
Grade 6			8	0.31	0.03
1	4.54	0.65	9	0.22	0.02
2	0.63	0.09	10	0.21	0.02
3	0.50	0.07	11	0.19	0.02
4	0.49	0.07			
5	0.37	0.05			
6	0.26	0.04			
7	0.21	0.03			

Table 7.9 Factor Loadings in Fixed-Number Factor Analysis with Orthogonal Varimax Rotation

Grade	Score Dimension		Factor 1	Factor 2	Factor 3	Factor 4	
3	Part I: Content Skills	Reading	0.29	0.22	0.88	0.21	
		Math	0.30	0.88	0.30	0.06	
	Part II: Generalization	Reading	0.35	0.11	0.51	0.73	
		Math	0.33	0.69	0.01	0.59	
	Part III: Opportunities	Modifications	0.82	0.21	0.15	0.25	
		Self Determinations	0.79	0.28	0.30	0.12	
		Interactions	0.69	0.28	0.28	0.39	
	4	Part I: Content Skills	Reading	0.28	0.21	0.20	0.87
			Writing	0.14	0.30	0.79	0.39
Math			0.11	0.87	0.21	0.28	
Part II: Generalization		Reading	0.57	0.15	0.18	0.60	
		Writing	0.44	0.22	0.80	0.06	
		Math	0.42	0.77	0.25	0.05	
Part III: Opportunities		Modifications	0.82	0.18	0.17	0.15	
		Self Determinations	0.70	0.20	0.21	0.40	
		Interactions	0.76	0.22	0.30	0.26	
5	Part I: Content Skills	Reading	0.27	0.17	0.24	0.88	
		Math	0.13	0.68	0.36	0.38	
		Science	0.24	0.15	0.88	0.29	
	Part II: Generalization	Reading	0.42	0.42	0.11	0.61	
		Math	0.32	0.88	0.17	0.13	
		Science	0.41	0.37	0.72	0.07	
	Part III: Opportunities	Modifications	0.85	0.17	0.19	0.18	
		Self Determinations	0.77	0.18	0.31	0.27	
		Interactions	0.76	0.32	0.25	0.28	
6	Part I: Content Skills	Reading	0.33	0.66	0.61	0.01	
		Math	0.26	0.15	0.84	0.38	
	Part II: Generalization	Reading	0.31	0.82	0.12	0.39	
		Math	0.32	0.26	0.30	0.82	
	Part III: Opportunities	Modifications	0.83	0.20	0.10	0.27	
		Interactions	0.69	0.33	0.28	0.36	
7	Part I: Content Skills	Reading	0.18	0.67	0.50	0.28	
		Writing	0.27	0.40	0.16	0.82	
		Math	0.38	0.82	0.12	0.28	
	Part II: Generalization	Reading	0.30	0.33	0.80	0.14	
		Writing	0.37	0.04	0.61	0.61	
		Math	0.45	0.59	0.42	0.09	
	Part III: Opportunities	Modifications	0.79	0.23	0.27	0.16	
		Self Determinations	0.71	0.38	0.12	0.37	
		Interactions	0.76	0.27	0.35	0.24	
8	Part I: Content Skills	Reading	0.23	0.90	0.20	0.10	
		Math	0.22	0.75	0.10	0.38	
		Science	0.30	0.32	0.84	-0.04	
	Part II: Generalization	Reading	0.26	0.57	0.25	0.49	
		Math	0.30	0.37	0.12	0.81	
		Science	0.32	0.05	0.79	0.73	
	Part III: Opportunities	Modifications	0.84	0.15	0.18	0.16	
		Self Determinations	0.79	0.28	0.31	0.16	
		Interactions	0.74	0.25	0.29	0.29	

Table 7.9 Factor Loadings in Fixed-Number Factor Analysis with Orthogonal Varimax Rotation
(Continued)

Grade	Score Dimension	Factor 1	Factor 2	Factor 3	Factor 4	
High School	Part I: Content Skills	Reading	0.67	0.19	0.30	0.29
		Writing	0.62	0.25	-0.01	0.68
		Math	0.78	0.27	0.28	0.10
		Science	0.79	0.20	0.34	0.08
	Part II: Generalization	Reading	0.21	0.22	0.74	0.28
		Writing	0.11	0.25	0.49	0.76
		Math	0.29	0.31	0.70	0.11
		Science	0.40	0.23	0.70	0.05
	Part III: Opportunities	Modifications	0.13	0.81	0.23	0.15
		Self Determinations	0.34	0.79	0.16	0.14
		Interactions	0.20	0.78	0.31	0.16

VIII. SUMMARY OF STUDENT PERFORMANCE

In this section, students’ performances are examined by analyzing the score distributions of all content areas and dimensions across all grade levels. Also, average scores are examined to provide information about how students did on WAAS.

The score distributions reported in Table 8.1 are based on the data captured in the demographic sheet and are the results reported to parents, schools, districts, and state. These data represent the final post-record results for the WAAS population. Table 8.1 provides a summary of the percentage of students earning each of the score points across four content areas and score dimensions. Unlike previous years, Part I scores are on a different scale from Part II scores. While Part I scoring remains on a 5-point scale (0, 1, 2, 3, 4), Part II contains only 3 score points (0, 1, 2). A major difference in the score distribution this year is that larger percentage of score zero can be observed across all content areas. To further investigate the score distribution, we dissect and analyze the distribution of portfolios that received a score of zero and find that significant portion of portfolios are identified as having insufficient evidence, which may be attributed to the new scoring rubrics and procedures implemented. Table 8.2 shows the score distribution based on the entire WAAS population including all grade levels. It can be seen that large percentages of portfolios are scored as having insufficient evidence (A), refusal (B), previously passed (C), or not tested (E). These scores (A, B, C, or E) are all converted to a score of zero.

Analysis of Part I Scores

In general, Part I scores follow a bimodal pattern where portfolios are either scored as 0 or 4. These two score points account for approximately 70% to 80% of the score distribution. If we exclude the zero-score portfolios and examine only the portfolios that are well aligned to GLE and receive a score other than zero, the majority of portfolios receive a score of 3 or 4, demonstrating good skills over various contents. In 2001, only 5.2% of the portfolios in Mathematics were scored at level 4. Percentages were low in the other domains as well. The distributions of the content scores gradually changed from year to year, with more students earning higher scores of 3 or 4. In 2006, the percentages of the highest score for all grades and contents ranged from 44.2% to 58.8%. This year, the percentages of the highest score range from 41.99% to 55.84%.

Analysis of Part II Scores

Scores on “Setting and Contexts” (Part II) also show good generalization of content skills. Similar to Part I scores, a bimodal distribution can be observed for all grades across all content areas. As indicated in Table 8.1 and Table 8.2, more portfolios receive a score of zero in Part II than in Part I. In other words, insufficient evidence in generalization of content skills is more common than insufficient evidence observed in content skills. However, if one excludes the zero-score cases and examines only the portfolios demonstrating good alignment to GLE and some level of generalization, most portfolios do successfully demonstrate skills in a variety of settings as the percentage of portfolios given a Part II score of 2 is consistently higher than the percentage of score of 1 for all grades across all content areas.

Total Score Distribution

With the observed bimodal distribution in both Part I and Part II scores, we can expect the total score (two times sum of Part I and Part II scores) to show a similar distributional pattern as well. Table 8.3 summarizes the distribution of total scores for all grades and contents. It shows the number and percentage of students at each score point.

Mean Score Summary

Student performance is examined by generating the average total scores for all content areas across all grades. This analysis is conducted at various levels. Table 8.4 presents the mean score summary based on all WAAS participants and grade-specific populations. Gender and ethnic subgroup analysis is also presented.

Table 8.1 2007 WAAS Score Distributions

Score Point	Part I: Content Skills				Part II: Generalization of Skills (Contexts and Setting)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Grade 3											
0	22.87	NA	26.36	NA	34.50	NA	37.73	NA	35.66	33.98	29.97
1	3.88	NA	3.75	NA	17.83	NA	15.76	NA	16.93	11.24	19.38
2	8.14	NA	8.01	NA	47.67	NA	46.51	NA	47.42	54.78	50.65
3	13.18	NA	12.79	NA	NA	NA	NA	NA	NA	NA	NA
4	51.94	NA	49.10	NA	NA	NA	NA	NA	NA	NA	NA
Grade 4											
0	24.59	27.64	27.26	NA	35.67	41.02	39.11	NA	39.75	35.41	33.63
1	3.95	4.08	4.46	NA	21.02	21.40	21.40	NA	21.02	14.65	21.02
2	9.55	7.90	6.88	NA	43.31	37.58	39.49	NA	39.24	49.94	45.35
3	10.32	16.94	11.85	NA	NA	NA	NA	NA	NA	NA	NA
4	51.59	43.44	49.55	NA	NA	NA	NA	NA	NA	NA	NA
Grade 5											
0	26.38	NA	26.52	36.19	39.78	NA	41.02	47.79	45.99	40.88	37.29
1	5.80	NA	4.97	4.83	18.51	NA	20.58	17.68	18.65	15.47	18.78
2	7.32	NA	7.04	5.39	41.71	NA	38.40	34.53	35.36	43.65	43.92
3	11.88	NA	12.29	11.60	NA	NA	NA	NA	NA	NA	NA
4	48.62	NA	49.17	41.99	NA	NA	NA	NA	NA	NA	NA

Values in cells represent percentages.

Table 8.1 2007 WAAS Score Distributions (Continued)

Score Point	Part I: Content Skills				Part II: Generalization of Skills (Contexts and Setting)				Part III: Student's Educational Opportunities		
	Reading	Writing	Math	Science	Reading	Writing	Math	Science	Modifications and Adaptations	Self Determination	Interactions
Grade 6											
0	22.40	NA	23.86	NA	37.18	NA	39.29	NA	41.88	39.61	35.55
1	5.19	NA	5.52	NA	19.81	NA	18.34	NA	16.23	10.71	16.72
2	8.12	NA	5.03	NA	43.02	NA	42.37	NA	41.88	49.68	47.73
3	8.44	NA	12.01	NA	NA	NA	NA	NA	NA	NA	NA
4	55.84	NA	53.57	NA	NA	NA	NA	NA	NA	NA	NA
Grade 7											
0	29.12	28.96	27.52	NA	43.36	46.08	41.60	NA	44.00	40.00	40.32
1	5.44	4.80	5.44	NA	16.00	18.88	14.56	NA	19.52	12.48	18.56
2	9.44	8.48	6.56	NA	40.64	35.04	43.84	NA	36.48	47.52	41.12
3	9.44	14.24	10.24	NA	NA	NA	NA	NA	NA	NA	NA
4	46.56	43.52	50.24	NA	NA	NA	NA	NA	NA	NA	NA
Grade 8											
0	30.73	NA	30.31	30.59	47.21	NA	46.09	44.97	47.49	40.50	37.57
1	4.61	NA	6.42	5.45	13.55	NA	14.11	14.11	15.78	10.89	17.04
2	7.12	NA	6.15	5.03	39.25	NA	39.80	40.92	36.73	48.60	45.39
3	11.59	NA	10.47	15.64	NA	NA	NA	NA	NA	NA	NA
4	45.95	NA	46.65	43.30	NA	NA	NA	NA	NA	NA	NA
High School											
0	27.80	27.13	28.13	38.65	44.52	47.73	46.18	55.04	45.51	38.32	35.55
1	4.32	4.43	4.54	4.10	16.17	17.83	18.05	13.40	19.60	13.84	24.81
2	6.64	7.53	7.31	5.76	39.31	34.44	35.77	31.56	34.88	47.84	39.65
3	12.40	13.62	12.18	12.18	NA	NA	NA	NA	NA	NA	NA
4	48.84	47.29	47.84	39.31	NA	NA	NA	NA	NA	NA	NA

Values in cells represent percentages.

Table 8.2 2007 WAAS Score Distribution Based on the Entire Population

score	Part I: Content Skills								Part II: Generalization of Skills (Contexts and Setting)								Part III: Student's Educational Opportunities					
	Reading		Writing		Math		Science		Reading		Writing		Math		Science		Modifications and Adaptations		Self Determination		Interactions	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
1	244	4.74	102	4.41	255	4.96	111	4.74	902	17.54	447	19.33	907	17.64	350	14.94	943	18.34	661	12.85	1016	19.76
2	411	7.99	183	7.91	349	6.79	127	5.42	2166	42.12	825	35.67	2091	40.66	828	35.34	1995	38.79	2515	48.90	2300	44.72
3	576	11.20	345	14.92	604	11.74	307	13.10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	2560	49.78	1040	44.96	2535	49.29	968	41.31	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
A	1041	20.24	487	21.05	987	19.19	530	22.62	1764	34.30	885	38.26	1732	33.68	865	36.92	2094	40.72	1856	36.09	1716	33.37
B	50	0.97	18	0.78	51	0.99	41	1.75	50	0.97	18	0.78	51	0.99	41	1.75	50	0.97	50	0.97	50	0.97
C	64	1.24	31	1.34	65	1.26	64	2.73	64	1.24	31	1.34	65	1.26	64	2.73	29	0.56	29	0.56	29	0.56
E	197	3.83	107	4.63	297	5.77	195	8.32	197	3.83	107	4.63	297	5.77	195	8.32	32	0.62	32	0.62	32	0.62

A=Insufficient Evidence; B=Refusal; C=Previously Passed; E=Not Tested

Table 8.3 2007 WAAS Total Score Distributions

Score	Reading		Writing		Math		Science	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Grade 3								
0	164	21.55	NA	NA	188	24.80	NA	NA
2	12	1.58	NA	NA	9	1.19	NA	NA
4	24	3.15	NA	NA	20	2.64	NA	NA
6	42	5.52	NA	NA	47	6.20	NA	NA
8	105	13.80	NA	NA	106	13.98	NA	NA
10	144	18.92	NA	NA	130	17.15	NA	NA
12	270	35.48	NA	NA	258	34.04	NA	NA
Grade 4								
0	188	24.10	206	26.68	206	26.51	NA	NA
2	9	1.15	13	1.68	9	1.16	NA	NA
4	28	3.59	23	2.98	21	2.70	NA	NA
6	54	6.92	69	8.94	56	7.21	NA	NA
8	98	12.56	105	13.60	105	13.51	NA	NA
10	148	18.97	157	20.34	153	19.69	NA	NA
12	255	32.69	199	25.78	227	29.21	NA	NA
Grade 5								
0	181	25.35	NA	NA	182	25.49	242	34.42
2	17	2.38	NA	NA	16	2.24	14	1.99
4	19	2.66	NA	NA	23	3.22	19	2.70
6	50	7.00	NA	NA	50	7.00	42	5.97
8	108	15.13	NA	NA	96	13.45	83	11.81
10	119	16.67	NA	NA	144	20.17	122	17.35
12	220	30.81	NA	NA	203	28.43	181	25.75

Table 8.3 2007 WAAS Total Score Distributions (Continued)

Score	Reading		Writing		Math		Science	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Grade 6								
0	127	20.99	NA	NA	134	22.22	NA	NA
2	11	1.82	NA	NA	19	3.15	NA	NA
4	28	4.63	NA	NA	14	2.32	NA	NA
6	40	6.61	NA	NA	27	4.48	NA	NA
8	84	13.88	NA	NA	91	15.09	NA	NA
10	97	16.03	NA	NA	127	21.06	NA	NA
12	218	36.03	NA	NA	191	31.67	NA	NA
Grade 7								
0	167	27.38	166	27.35	156	25.62	NA	NA
2	19	3.11	13	2.14	18	2.96	NA	NA
4	23	3.77	26	4.28	21	3.45	NA	NA
6	42	6.89	58	9.56	40	6.57	NA	NA
8	77	12.62	84	13.84	66	10.84	NA	NA
10	90	14.75	103	16.97	89	14.61	NA	NA
12	192	31.48	157	25.86	219	35.96	NA	NA
Grade 8								
0	170	25.53	NA	NA	172	25.63	167	25.15
2	22	3.30	NA	NA	28	4.17	16	2.41
4	21	3.15	NA	NA	16	2.38	17	2.56
6	42	6.31	NA	NA	30	4.47	45	6.78
8	91	13.66	NA	NA	111	16.54	111	16.72
10	115	17.27	NA	NA	112	16.69	103	15.51
12	205	30.78	NA	NA	202	30.10	205	30.87
High School								
0	189	22.53	189	22.34	189	22.58	196	26.17
2	13	1.55	22	2.60	22	2.63	12	1.60
4	28	3.34	36	4.26	29	3.46	26	3.47
6	61	7.27	54	6.38	65	7.77	74	9.88
8	152	18.12	172	20.33	136	16.25	123	16.42
10	134	15.97	149	17.61	162	19.35	107	14.29
12	262	31.23	224	26.48	234	27.96	211	28.17

Table 8.4 Mean Total Score Summary for All Grades and Contents

Group	Count	Mean Total Score Reading	Mean Total Score Writing	Mean Total Score Math	Mean Total Score Science
Grade 3					
All	774	7.74	NA	7.42	NA
Male	520	7.83	NA	7.48	NA
Female	254	7.57	NA	7.29	NA
Native American	27	9.38	NA	6.89	NA
Asian	51	8.24	NA	9.10	NA
African American	65	7.26	NA	6.74	NA
Hispanic	99	7.78	NA	7.56	NA
White	494	7.69	NA	7.37	NA
Grade 4					
All	785	7.41	6.90	7.12	NA
Male	532	7.33	7.05	7.08	NA
Female	253	7.58	6.60	7.20	NA
Native American	36	7.44	7.89	8.17	NA
Asian	57	6.77	6.18	7.02	NA
African American	63	6.98	7.08	7.68	NA
Hispanic	129	7.43	6.89	7.24	NA
White	469	7.51	6.85	6.89	NA
Grade 5					
All	724	7.15	NA	7.10	6.28
Male	468	7.23	NA	7.05	6.35
Female	256	7.00	NA	7.18	6.13
Native American	23	8.35	NA	7.83	6.27
Asian	46	6.13	NA	6.61	6.26
African American	52	7.71	NA	7.63	6.98
Hispanic	116	7.65	NA	7.70	6.61
White	459	7.09	NA	6.98	6.13

Statistics for Pacific Islander and Multi Racial students not reported due to their size (less than 20 in total).

Table 8.4 Mean Total Score Summary for All Grades and Contents (Continued)

Group	Count	Mean Total Score Reading	Mean Total Score Writing	Mean Total Score Math	Mean Total Score Science
Grade 6					
All	616	7.66	NA	7.54	NA
Male	394	7.50	NA	7.40	NA
Female	222	7.93	NA	7.78	NA
Native American	24	7.92	NA	7.25	NA
Asian	47	7.87	NA	7.66	NA
African American	42	8.52	NA	7.43	NA
Hispanic	68	8.03	NA	8.00	NA
White	411	7.43	NA	7.51	NA
Grade 7					
All	625	6.89	6.70	7.23	NA
Male	406	6.78	6.68	7.18	NA
Female	219	7.09	6.73	7.33	NA
Native American	22	5.27	5.09	6.18	NA
Asian	45	6.09	5.51	6.98	NA
African American	44	7.07	6.88	6.14	NA
Hispanic	83	6.59	5.90	7.59	NA
White	403	7.12	7.04	7.39	NA
Grade 8					
All	716	7.08	NA	7.05	7.15
Male	438	7.11	NA	7.00	7.37
Female	278	7.05	NA	7.13	6.80
Native American	21	5.80	NA	8.40	5.52
Asian	44	6.14	NA	6.47	5.40
African American	56	8.04	NA	6.92	7.65
Hispanic	91	6.69	NA	6.79	6.46
White	470	7.20	NA	7.05	7.42

Statistics for Pacific Islander and Multi Racial students not reported due to their size (less than 20 in total).

Table 8.4 Mean Total Score Summary for All Grades and Contents (Continued)

Group	Count	Mean Total Score Reading	Mean Total Score Writing	Mean Total Score Math	Mean Total Score Science
High School					
All	902	7.39	7.17	7.24	6.89
Male	541	7.46	7.18	7.26	7.09
Female	361	7.28	7.15	7.22	6.58
Asian	26	5.85	5.46	6.54	5.62
African American	41	7.33	7.19	7.55	7.21
Hispanic	92	7.40	7.72	7.70	7.63
White	468	7.62	7.25	7.30	7.11

Statistics for Native American, Pacific Islander, and Multi Racial students not reported due to their size (less than 20 in total).

IX. PERFORMANCE ON THE WAAS RELATIVE TO STANDARDS

Federal legislation and regulations for ESEA and IDEA reauthorization require states to report results for all students assessed using general assessments and alternate assessments relative to the same grade level academic content and achievement standards. The Office of Superintendent of Public Instruction reestablished four levels of performance based on alternate achievement standards on the WAAS assessments in the summer of 2007.

Educators from throughout the State of Washington met in Bellevue on June 25-28, 2007, to recommend cut scores for the Washington Alternate Assessment System (WAAS) Portfolio Assessment. Measurement Incorporated staff assisted the Office of Superintendent of Public Instruction conduct the meeting. A total of 42 Washington educators (standard-setting panelists) participated in the process which consisted of a generalized holistic review of amalgamated student portfolios (so called because each portfolio consisted of elements from portfolios of two different students).

The standards for the portfolio were set using performance-level descriptors (Appendix C) and copies of portfolios. Panelists received a thorough orientation to the scope and purpose of the WAAS, its scoring rubrics, and its role in the State's accountability system. Panelists also received a thorough introduction to the generalized holistic standard-setting procedure, as well as an opportunity to practice using the procedure prior to actual standard setting. These panelists were divided into four groups of nine to fourteen people:

Grades 3-4	9 panelists
Grades 5-6	10 panelists
Grades 7-8	9 panelists
Grade 10	14 panelists

Within each panel, members evaluated portfolios for all grade-cluster content areas: reading and mathematics in all grades, writing in grades 4, 7, and 10; and science in grades 5, 8, and 10.

Facilitators from Measurement Incorporated led panelists through three rounds of review and discussion of the portfolios. In Round 1, panelists reviewed 17 to 19 separate portfolios,

classifying them into one of four performance categories: Well Below Standard (Level 1), Approaches Standard (Level 2), Meets Standard (Level 3), and Exceeds Standard (Level 4). The facilitators entered panelists' ratings and then used the relationship between these ratings and the scores previously assigned to the portfolios to establish cut scores. Panelists discussed the results of Round 1, noting differences in their application of performance level descriptors (PLDs). After this discussion, panelists again rated the portfolios (classifying each at one of the four performance levels). After Round 2, facilitators again entered and analyzed the ratings and presented the results to the panelists, along with student score distributions. This information allowed panelists to see the impact of their deliberations on students throughout Washington. The panelists considered all this information and ultimately entered three cut scores, one each for Approaches Standard, Meets Standard, and Exceeds Standard. Facilitators calculated the means of each of these sets of cut scores.

Following the Round 3 recommendations, representatives from each panel met to resolve any cut score differences across grade clusters. This meta-panel was referred to as the Articulation Committee, so-called because their charge was to make sure that cut scores from one grade cluster to the next did not result in wild swings of numbers or percentages of students being classified at any performance level. Table 9.1 shows all cut scores and the resulting percentages of students classified in each category after Round 3. Table 9.2 shows the resulting cut scores and percentages of students classified in each category after articulation. The shaded portions of Table 9.2 show the four changes recommended by the Articulation Committee. Each change was by a single point.

Table 9.1 Cut Scores and Percentages of Students in Each Category after Round 3

Grade/Subject	Level 1: Well Below Standard		Level 2: Approaches Standard		Level 3: Meets Standard		Level 4: Exceeds Standard		Level 3 & 4 Meets + Exceeds	
	Cut	%	Cut	%	Cut	%	Cut	%	Cut	%
3-4 Reading	---	23	3	7	6	16	9	54	6	71
3-4 Mathematics	---	21	3	6	6	18	9	54	6	72
4 Writing	---	23	3	3	5	17	8	57	5	74
5-6 Reading	---	25	4	9	7	23	10	43	7	66
5-6 Mathematics	---	23	3	6	6	18	9	53	6	71
5 Science	---	32	4	8	7	22	10	37	7	59
7-8 Reading	---	27	4	9	7	23	10	43	7	64
7-8 Mathematics	---	27	4	7	7	23	10	43	7	66
7 Writing	---	29	4	12	7	23	10	36	7	59
8 Science	---	27	4	8	7	25	10	40	7	65
HS Reading	---	20	3	6	6	27	10	47	6	74
HS Mathematics	---	24	3	4	6	23	10	49	6	72
HS Writing	---	23	3	5	6	27	10	45	6	72
HS Science	---	25	4	10	7	21	10	44	7	65

Table 9.2 Cut Scores and Percentages of Students in Each Category after Articulation

Grade/Subject	Level 1: Well Below Standard		Level 2: Approaches Standard		Level 3: Meets Standard		Level 4: Exceeds Standard		Level 3 & 4 Meets + Exceeds	
	Cut	%	Cut	%	Cut	%	Cut	%	Cut	%
3-4 Reading	---	23	3	7	6	16	9	54	6	71
3-4 Mathematics	---	21	3	6	6	18	9	54	6	72
4 Writing	---	23	3	3	6	17	9	57	6	74
5-6 Reading	---	25	4	9	7	23	10	43	7	66
5-6 Mathematics	---	23	3	6	6	18	9	53	6	71
5 Science	---	32	4	8	7	22	10	37	7	59
7-8 Reading	---	27	4	9	7	23	10	43	7	64
7-8 Mathematics	---	26	3	6	6	25	10	43	6	68
7 Writing	---	29	4	12	7	23	10	36	7	59
8 Science	---	27	4	8	7	25	10	40	7	65
HS Reading	---	20	3	6	6	27	10	47	6	74
HS Mathematics	---	24	3	4	6	23	10	49	6	72
HS Writing	---	23	3	5	6	27	10	45	6	72
HS Science	---	25	4	10	7	21	10	44	7	65

Table 9.3 shows the actual percentage of students achieving standards on the 2007 WAAS for each grade and content area. The achievement standards reported here are for the WAAS assessments and should not be compared to the results or standards for students taking the WASL.

Table 9.3 Summary of Distribution by Performance Level on 2007 WAAS

		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	HS
Reading								
N count		774	785	724	616	625	716	903
Level 1	Not Meeting Standard	23.13%	25.26%	27.73%	22.81%	30.49%	28.83%	24.08%
Level 2		3.15%	3.59%	9.66%	11.24%	10.66%	9.46%	3.34%
Level 3	Meeting Standard	19.32%	19.49%	15.13%	13.88%	12.62%	13.66%	25.39%
Level 4		54.40%	51.67%	47.48%	52.07%	46.23%	48.05%	47.20%
Writing								
N count			785			625		903
Level 1	Not Meeting Standard		28.37%			23.49%		24.94%
Level 2		2.98%				13.84%		4.26%
Level 3	Meeting Standard		22.54%			13.84%		26.71%
Level 4		46.11%				42.83%		44.09%
Math								
N count		774	785	724	616	625	716	903
Level 1	Not Meeting Standard	25.99%	27.67%	27.73%	25.37%	28.57%	29.81%	25.21%
Level 2		2.64%	2.70%	3.22%	2.32%	3.45%	2.38%	3.46%
Level 3	Meeting Standard	20.18%	20.72%	20.45%	19.57%	17.41%	21.01%	24.01%
Level 4		51.19%	48.91%	48.60%	52.74%	50.57%	46.80%	47.31%
Science								
N count				724			716	903
Level 1	Not Meeting Standard			36.42%			27.56%	27.77%
Level 2		8.68%					9.34%	13.35%
Level 3	Meeting Standard			11.81%			16.72%	16.42%
Level 4		43.10%					46.39%	42.46%

Performance distributions were also examined separately for males and females. Table 9.4 shows the percentage of students classified into each performance level for both genders across all grades. In general, the percentages of students meeting the new standard are very close for males and females. Girls generally have a higher percentage of level-4 performance than boys at middle and high schools except on high school math and science. Also, girls seem to perform slightly better than boys at grade 6. Differences are all pretty minor at other grade levels and content areas.

Table 9.4 2007 WAAS Performance Distribution by Gender

		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		HS	
Reading		M	F	M	F	M	F	M	F	M	F	M	F	M	F
N count		520	254	532	253	468	256	394	222	406	219	438	278	541	361
Level 1	Not Meeting Standard	22.64	24.11	25.33	25.10	27.06	28.97	23.58	21.46	30.23	30.99	28.54	29.30	22.80	26.04
Level 2		2.76	3.95	3.97	2.79	9.09	10.71	13.47	7.31	12.59	7.04	9.27	9.77	3.80	2.66
Level 3	Meeting Standard	19.29	19.37	20.23	17.93	15.58	14.29	11.92	17.35	13.10	11.74	14.63	12.11	26.80	23.37
Level 4		55.31	52.57	50.47	54.18	48.27	46.03	51.04	53.88	44.08	50.23	47.56	48.83	46.60	47.93
Writing		M	F	M	F	M	F	M	F	M	F	M	F	M	F
N count				532	253					406	219			541	361
Level 1	Not Meeting Standard			26.86	31.58					28.10	32.08			24.41	25.82
Level 2				2.86	3.24					15.95	9.91			4.53	3.86
Level 3	Meeting Standard			24.00	19.43					14.43	12.74			27.95	24.63
Level 4				46.29	45.75					41.52	45.28			43.11	45.70
Math		M	F	M	F	M	F	M	F	M	F	M	F	M	F
N count		520	254	532	253	468	256	394	222	406	219	438	278	541	361
Level 1	Not Meeting Standard	24.90	28.17	27.76	27.49	28.94	25.50	26.42	23.50	28.28	29.11	29.23	30.74	26.52	23.39
Level 2		2.96	1.98	3.04	1.99	1.94	5.58	2.33	2.30	3.03	4.23	2.17	2.72	3.24	3.80
Level 3	Meeting Standard	21.15	18.25	20.53	21.12	20.30	20.72	19.43	19.82	19.19	14.08	22.71	18.29	21.66	27.49
Level 4		50.99	51.59	48.67	49.40	48.81	48.21	51.81	54.38	49.49	52.58	45.89	48.25	48.58	45.32
Science		M	F	M	F	M	F	M	F	M	F	M	F	M	F
N count						468	256					438	278	541	361
Level 1	Not Meeting Standard					35.67	37.80					25.12	31.50	26.35	29.93
Level 2						7.88	10.16					9.27	9.45	13.06	13.82
Level 3	Meeting Standard					13.13	9.35					19.27	12.60	15.77	17.11
Level 4						43.33	42.68					46.34	46.46	44.82	39.14

APPENDIX A. WAAS PORTFOLIO SCORING 2007 SCORING HANDBOOK

General Review of the Portfolio

Prior to scoring the portfolio:

- Review the student information form to make sure all information on the binder label matches the information sheet.
 - Make sure student name matches the binder label
 - Make sure SSID is filled in
 - **Make sure the grade level matches the binder label******
- If any of the information on the binder label does not match the student information sheet, see your table leader.

General Review of Each Entry

Prior to scoring each entry:

1. Review the entry cover sheet and evidence and verify:
 - Grade of student listed
 - EALR & GLE are listed
 - Targeted skill is listed
 - Measurable criterion is linked to the targeted skill
2. Review each entry to ensure that it contains 5 or more pieces of evidence. If there are not 5 pieces, bring to your table leader.
3. If a content area contains 2 entries, consult your table leader.

Explanation of Portfolio Scoring Criteria

The following definitions and clarifications of the scoring rubric are given as guides in assigning the correct scores to each portfolio. These clarifications are derived from the key language in the scoring rubric and are evidenced in the exemplars provided in training. The rubric has three parts. **Part one** assesses evidence as it relates to the targeted skill and performance of that targeted skill for **each entry** in each content area. **Part two** assesses the extent to which the student can generalize the targeted skill **across each content area**. **Part three** assesses the educational opportunities provided to the student **across all content areas**.

Part 1: Evidence of Targeted Skills

Evidence Aligned to the Targeted Skill

Scores	4	3	2	1	IE
Evidence Aligned to Targeted Skill	All evidence is aligned to the targeted skill	Most evidence is aligned to the targeted skill	Some evidence is aligned to the targeted skill	Very little evidence aligned to the targeted skill	No evidence aligned to the targeted skill

Definitions

Evidence: any type of direct observation, product, or constructed response related to the performance of the targeted skill, such as student work samples, data collection (i.e. charts, graphs, etc.), notes from parents, peers, general ed. teachers, and/or video or audio tapes.

Alignment: the focus of the evidence maintains fidelity with the content of the targeted skill (i.e. Does the evidence reflect the performance/behavior/content described in the targeted skill?)

Targeted Skill: specific measurable behavior in the content area to be assessed

Scoring Clarifications for Evidence Aligned to Targeted Skill

1. Always refer to the entry cover sheet for the targeted skill
2. If the entry cover sheet does not contain the GLE or Targeted Skill or there is no entry cover sheet, consult your table leader.
3. Review the evidence and verify that the student's performance is based on work that is representative of the targeted skill.
4. If there is a discrepancy between the targeted skill written on the entry cover sheet and the targeted skill written on the evidence, consult your table leader.
5. To reach a score point of four (4) **ALL** evidence within the content area must align to / reflect the targeted skill.
6. To reach a score point of three (3), **most** evidence within the content area must align to / reflect the targeted skill. Most refers to more than half of the evidence.
7. To reach a score point of two (2), **some** evidence within the content area must align to / reflect the targeted skill. Some refers to less than half.
8. To receive a score point of one (1), **very little** evidence within the content area must align to / reflect the targeted skill. Very little refers to one piece of evidence.
9. To receive a score point of Insufficient Evidence (IE), **NO** evidence within the content area alignes to / reflects the targeted skill.
10. If an entry scores IE for Evidence to Targeted Skill, score IE for Evidence of Performance and for Settings and Contexts. Do not use the entry when scoring the Educational Opportunities.

Scorer Notes

Evidence of Performance on the Targeted Skill

	4	3	2	1	IE
Evidence of performance on targeted skill	Evidence demonstrates that student exceeded the goal for the targeted skill.	Evidence demonstrates that student met the goal for the targeted skill.	Evidence demonstrates that student approached the goal for the targeted skill.	Evidence demonstrates that the student is not approaching the goal on the targeted skill.	No evidence of student's level of proficiency on the targeted skill

Definitions

Performance must be evidenced by:

- Data collection points (student data sheets)
 - Charts
 - Graphs

Performance may be backed up by:

- Teacher observation data and/or anecdotal statements accompanying a piece of evidence
- Student work samples
- Photographs, videotapes, or audiotapes that show data, growth achievement, and performance of the skill

Prior to Scoring for Performance

1. Review the evidence and verify that the student's name and date (month/day/year) are on each piece of evidence.
 - If the name is missing, **DO NOT use this piece** of evidence when you are scoring the entry.
 - If a date is missing, **DO NOT use this piece** of evidence when you are scoring the entry.
 - Exception: if the month and year are listed for August through March, but no day, this evidence may still be used; April must have month/day/year.
2. Review evidence and verify that dates appearing on the evidence are within the timeframe of August, 2006 through April 6, 2007. (The March 30 end date was extended due to snow.)
3. If the dates on any evidence are earlier or later than this timeframe, **DO NOT use** when scoring the entry.

Scoring Clarifications for Performance

1. Always refer to the entry cover sheet for the targeted skill and criterion.
2. If the targeted skill is not observable or measurable, consult your table leader
3. If no criterion is stated, 100% is assumed to be the goal.
4. Only use evidence that is based on work that is representative of the targeted skill as determined in the previous dimension (Evidence Aligned to Targeted Skill).
5. Data must be included as evidence of performance against the criterion. **If a data chart is not evidenced within the entry, the entry will score IE.** Data can be data sheets or graphs and must include
 - Student's name
 - Dates of data collection including month, day and year
 - Targeted skill being measured
 - Key (if needed)
 - Graphs must have the vertical and horizontal axis clearly labeled.
6. Review any data charts or graphs.
 - Review the row and column titles, any legends that appear on the chart, and the data that fill the chart. If you cannot interpret the legend or the key, consult your table leader.
 - A data chart or graph must include at least three discrete data collection points. If there are only one or two points, the performance dimension will score a one (1) "...not approaching the goal..."

- If you spend more than 5 minutes trying to interpret the evidence, consult your table leader.

7. Data must clearly indicate performance independent of a note or declaration of “met or exceeded goal”.

Scoring Clarifications for Performance continued

8. If a student starts at or above the criterion, the entry will score a one (1) for performance, unless the teacher increases the goal, changes the target, or continues to challenge the student in another way, then it may be scored *based on the new goal*.

9. If the word “trial” is used in the criterion and it is not defined, then it is assumed a trial refers to one data date.

10. If an entry scores IE for Evidence of Performance, the entry will score IE for Settings and Contexts and cannot be used for scoring the Educational Opportunities.

11. If the criterion contains 2 parts and the student meets criterion on one part and exceeds it on another, it scores a 4:

- For Example: 95% for 3 out of 4 sessions
 - Data is 90%, 96%, 98%, 96% - met 3 out of 4 sessions and exceeded 95%
 - Data is 95%, 95%, 95%, 95% - met 95% and exceeded 3 out of 4 sessions

12. If the targeted skill contains multiple goals and the evidence documents multiple goals, consult your table leader for scoring clarification.

- For Example: The student will increase reading vocabulary by 20 words and read grade level passages with 80% accuracy for 3 out of 5 trials and will read a new passage at 40 words per minute.

Scorer Notes:

Student Generalization of Skills (This dimension is scored for each content area)

Settings and Contexts

	2	1	IE
Setting and Contexts	Evidence that student demonstrates skill(s) in a variety of settings.	Evidence that student demonstrates skill(s) in two settings.	Evidence that student demonstrates skill(s) in only one setting.

Definitions

Settings: instructional environments where the learning, practice, and demonstration of skills occurs

Contexts: application of the targeted skill in different ways and activities

Scoring Clarifications for Settings and Context

- **A variety of settings or contexts** refers to multiple settings in and out of school and/or activities in which performance of the targeted skill occurs

Settings:

1. Settings are different locations in which the student demonstrates the targeted skill
2. Settings must have an authentic link to the skill and contexts
3. School settings
4. Community/Home
5. A setting must be clearly documented

Contexts:

1. Number of different activities/applications in which the student demonstrates the targeted skill
2. Contexts must have an authentic link to the skill and setting

Contexts occur within settings and therefore would score as one setting/context (i.e. playing a money game in the classroom counts as one; purchasing an item in a store counts as a second).

1. Different content classes may be considered different contexts even if the setting remains the same. For example, elementary aged students may spend their entire day in the same classroom while demonstrating skills in a variety of different content classes (i.e. a student may be learning decoding skills in reading class and applying those skills in science class to read new science vocabulary. This would evidence two different settings.)
2. Referencing different instructional groupings within a content class (i.e. 1:1 instruction; small group instruction; or large class instruction) would evidence different contexts in which the student practices the skill even though the setting remains the classroom

When scoring for settings and contexts:

- A content area must evidence 3 or more settings and/or contexts to score a two (2)
- A content area must evidence 2 settings and/or contexts to score a one (1)
- A content area that contains 1 or no settings and/or contexts will score Insufficient Evidence (IE)

Scorer Notes for Settings and Contexts:

Part III: Student’s Educational Opportunities (This dimension is scored across *all content areas*)

Modifications and Adaptations

	2	1	IE
Modifications And Adaptations	Multiple sources of evidence show that the student uses appropriate supports, adaptations and/or assistive technology in demonstrating skill(s) <i>in the content area</i> .	Some evidence shows that the student uses appropriate supports, adaptations and/or assistive technology in demonstrating skill(s) <i>in the content area</i> .	Little or no evidence shows that the student uses appropriate supports, adaptations and/or assistive technology in demonstrating skill(s) <i>in the content area</i> .

Support: the degree to which the adaptations, modifications, or assistive technology enabled the student to perform more independently

Modifications: substantial changes from the grade-level curriculum related to what a student is expected to learn and/or demonstrate

Adaptations: alterations made to existing materials or instructional delivery in order to make the curriculum accessible to the student

Assistive Technology: devices/equipment used to assist students in developing and participating in meaningful standards-based instruction

Independence: student performances that are done spontaneously without cues and prompts other than those that occur naturally (i.e., “authentic”)

Natural Support: support provided by others in the context of students learning together (e.g., partners in a science experiment, co-workers in a job placement)

Scoring Clarifications for Modifications and Adaptations:

1. Adaptations/modifications/assistive technology are not differentiated for scoring purposes.
2. Adaptations/modifications/assistive technology must be used within the performance of the targeted skill to be considered in scoring.
3. Prompts are not adaptations, modifications or assistive technology. Prompts are people dependent; they are not supports that lead to independence. Use of prompts always requires direct assistance from other people. However, other people assigned to assist the student (even scribes and readers) should not be thought of as adaptations since they do not allow students to perform independently.
4. Specialized curricula are not considered adaptations/modifications/or assistive technology (Edmark Reading, Dolch lists, Touch Math).
5. Classroom materials needed to implement the activity for any child to use is not considered a modification or support (i.e. bingo game, flashcards, standard use of the computer, lower grade level materials, etc.).

Modifications and Adaptations are scored across the entire portfolio.

When a portfolio contains 3 or 4 different content entries, score as follows:

- (2) Score a two for **Multiple** sources of evidence that show the student using appropriate supports, supports must be evidenced within every content area.
- (1) Score a one for **Some** evidence that shows the student using appropriate supports, supports must be evidenced in most content areas.
- (IE) Score Insufficient Evidence for **little or no** evidence that shows the student using appropriate supports, supports may be evidenced only one time in one content area, or not evidenced at all.

When a portfolio contains 2 different content entries, score as follows:

- (2) Score a two when student use of appropriate supports is evidenced in both content areas.
- (1) Score a one when student use of appropriate supports is evidenced in one content area.
- (IE) Score Insufficient Evidence when student use of appropriate supports is not evidenced at all.

When a portfolio contains only 1 content entry, score as follows:

- (2) Score a two when student use of appropriate supports is evidenced in the content area.
- (IE) Score Insufficient Evidence when student use of appropriate supports is not evidenced at all.

Scorer Notes:

Self-Determination

	2	1	IE
Self Determination	Multiple and varied pieces of evidence show that student has made choices and contributed to planning, monitoring, and/or evaluating own skill development.	Several pieces of evidence show that the student has made choices and contributed to planning, monitoring, and/or evaluating own skill development.	Little or no evidence shows that the student has made choices and contributed to planning, monitoring, and/or evaluating own skill development.

Self-determination – the ability to control basic decisions and directions of one’s life and the level of student involvement in IEP educational activities; the degree to which a student participates in monitoring, evaluating and goal setting activities for targeted skills.

Monitoring – student maintains a record of performance of activity or of steps within the activity (e.g., identifies correct and incorrect work; checks off tasks performed at work or in school, checks off weekly scheduled tasks performed correctly at home, etc.); monitoring must be done by the student and should relate to the targeted skill/IEP objective specified on the entry cover sheet.

Evaluating – student reviews his/her own performance and /or identifies the quality of own performance and the components involved in that performance; evaluation must be an integral part of daily instruction, not occurring only at the end of an entry; evaluation must be done by the student and must relate to the targeted skill specified on the entry cover sheet.

Sets Goals (planning) – student makes decisions based upon evaluations regarding how to improve his/her performance (examples of goal-setting include planning what strategy to use, determining how to organize a task , or deciding how much work to do)

***Note:** This dimension measures a student’s participation in making choices **related to planning, monitoring, and evaluating** his or her work as it applies to the targeted skill.

Scoring Clarifications for Self Determination

1. Self determination can be evidenced through choice making as it relates to planning, monitoring or evaluating skill, product, or performance on the skill. Examples may include
 - Planning: The student may be involved in setting up the activity, planning what skills to practice, planning strategies for improvement, or planning how to practice the targeted skill, etc. This is usually done prior to the start of the next activity.
 - Monitoring: The student may be monitoring the steps of the activity, the completion of parts of the activity or charting progress on reaching the criteria on the IEP skill, etc.
 - Evaluation: The student may be evaluating the level of success, performance, etc. on the skill used during the activity.
2. General student comments, such as liking an activity or working hard, are not considered self-determination. Evaluation refers to an appraisal of performance, not a preference or emotional response.
3. A student choosing to participate in or complete the activity is not considered self-determination.
4. Making choices may not indicate self determination. When you see the word “choose” or “choice”, determine if the choices relate to planning, monitoring, and/or evaluating work on the targeted skill.
5. The student does not have to participate in self determination independently.

Self Determination is scored across the entire portfolio.

When a portfolio contains 3 or 4 different content entries, score as follows:

(2) Score a two when student use of Self Determination is evidenced within every content area.

(1) Score a one when student use of Self Determination is evidenced in most content areas.

(IE) Score Insufficient Evidence when student use of Self Determination is evidenced only one time in one content area, or not evidenced at all.

When a portfolio contains 2 different content entries, score as follows:

(2) Score a two when student use of Self Determination is evidenced in both content areas.

(1) Score a one when student use of Self Determination is evidenced in one content area.

(IE) Score Insufficient Evidence when student use of Self Determination is not evidenced at all.

When a portfolio contains only 1 content entry, score as follows:

(2) Score a two when student use of Self Determination is evidenced in the content area.

(IE) Score Insufficient Evidence when student use of Self Determination is not evidenced at all.

Interactions

Social Relationships / Interactions	2	1	IE
	Evidence shows that student has varied interactions with others related to the academic skills and/or in work, social, and/or community situations.	Evidence shows that student has some interactions with others related to the academic skills and/or in work, social, and/or community situations.	Little or no evidence that student has interactions with others related to academic skills and/or in work, social, and/or community situations.

Social Relationships/Interactions – interactions during targeted skill activities involving other students and adults, with and without disabilities, leading to the development of appropriate academic, motor, communication and social skills

Varied Interactions – interactions that occur across different settings and contexts (individual, group, leisure, academic, etc.)

Scoring Clarifications for Social Relationships/Interactions

1. For this dimension, the variety of social relationships/interactions between the student and others, both with and without disabilities, is examined as a community of learners.
2. A list that identifies a peer by name is not enough information to be considered when scoring for social relationships/interactions. More information is needed to determine if the student is working within the context of the targeted skill as a community of learners through a piece of acceptable evidence (i.e. student work sample, photo with caption, peer note, etc.). A description may accompany the evidence to further clarify the interaction.
3. If all social relationships/interactions evidenced occur with classmates or are labeled only as peers it will not score higher than a one.
4. Social relationships/interactions with family or staff paid to work directly with the student (teacher, paraprofessional, OT/PT/SLP) do not count as evidence of social interaction.
5. Social relationships/interactions with school office, cafeteria, or custodial staff, or with other community members (i.e., store clerk, postal employees, coworkers that work for a community business) can be used as evidence of interactions as long as the student is working on the targeted skill.
6. A photo of the student sitting with a group of peers in the library, captioned in general terms (“Student sitting with classmates in the library”) does not carry evidence of any meaningful interactions in itself. Further documentation of the quality and/or duration of those interactions would be necessary either from the student, peers, or an anecdotal comment from the teacher.

Social Relationships/Interactions are scored across the entire portfolio.

When a portfolio contains 3 or 4 different content entries, score as follows:

(2) Score a two when the evidence indicates varied interactions with others related to the targeted skills in **all** content areas.

(1) Score a one when the evidence indicates interactions with others related to the targeted skills in **most** content areas.

(IE) Score Insufficient Evidence when the evidence indicates varied interactions with others related to the targeted skills in one content area, or not evidenced at all.

When a portfolio contains 2 different content entries, score as follows:

(2) Score a two when the varied interactions with others related to the targeted skills are evidenced in both content areas.

(1) Score a one when interactions with others related to the targeted skills are evidenced in one content area.

(IE) Score Insufficient Evidence when interactions with others are not evidenced at all.

When a portfolio contains only 1 content entry, score as follows:

(2) Score a two when varied interactions with others related to the targeted skills are evidenced in the content area.

(IE) Score Insufficient Evidence when interactions are not evidenced at all.

Table Leader Procedures

1. Back read at least one scored portfolio for each scorer each day. Scorers should put their first scored portfolio of the day in the box labeled, "To Be Back Read." Your score will be a second read.
2. Note and retrain scorers at your table if they are discrepant with your scores.
3. Complete the second Scoring Monitor.
4. Return the scored portfolio to the box labeled "To be filed." Put the second score monitor in the tray.
5. Assist scorers when they have questions.
6. Attend table leader meetings and bring any concern to the attention of the leadership team.

APPENDIX B. WAAS PORTFOLIO SCORING SUMMARY SHEET—2007

Part I: Evidence of Targeted Skills (These rubrics were used to evaluate the evidence for *each skill* within each content area.)

Scores	4	3	2	1	IE
Evidence Aligned to Targeted Skill	All evidence is aligned to the targeted skill	Most evidence is aligned to the targeted skill	Some evidence is aligned to the targeted skill	Very little evidence aligned to the targeted skill	No evidence aligned to the targeted skill
Evidence of performance on targeted skill	Evidence demonstrates that student exceeded the goal for the targeted skill.	Evidence demonstrates that student met the goal for the targeted skill.	Evidence demonstrates that student approached the goal for the targeted skill.	Evidence demonstrates that the student is not approaching the goal on the targeted skill.	No evidence of student's level of proficiency on the targeted skill

CONTENT AREA	Skill Aligned to GLE (Y/N)	Evidence is Aligned to Skill (IE, 1-4)
Reading		
Writing		
Mathematics		
Science		

SKILL Score (IE, 1-4)

Part II: Generalization of Content Skills (This dimension is scored for each content area)

	2	1	IE
Setting and Contexts	Evidence that student demonstrates skill(s) in a variety of settings.	Evidence that student demonstrates skill(s) in two settings.	Evidence that student demonstrates skill(s) in only one setting.

Content Area Total Scores

	Part I Total (IE, 1-4)	Part II Total (IE, 1-2)	Part I + Part II Total Scores (IE, 2-6)
Reading (Grade 3-8,HS)			
Writing (Grade 4,7,HS)			
Mathematics (Grade 3-8,HS)			
Science (Grade 5,8,HS)			

Part III: Student's Educational Opportunities (This dimension is scored across *all content areas*)

	2	1	IE	Part III Scores
Modifications And Adaptations	Multiple sources of evidence show that the student uses appropriate supports, adaptations and/or assistive technology in demonstrating skill(s) <i>in the content area</i> .	Some evidence shows that the student uses appropriate supports, adaptations and/or assistive technology in demonstrating skill(s) <i>in the content area</i> .	Little or no evidence shows that the student uses appropriate supports, adaptations and/or assistive technology in demonstrating skill(s) <i>in the content area</i> .	
Self Determination	Multiple and varied pieces of evidence show that student has made choices and contributed to planning, monitoring, and/or evaluating own skill development.	Several pieces of evidence show that the student has made choices and contributed to planning, monitoring, and/or evaluating own skill development.	Little or no evidence shows that the student has made choices and contributed to planning, monitoring, and/or evaluating own skill development.	
Interactions	Evidence shows that student has a varied interactions with others related to the academic skills and/or in work, social, and/or community situations.	Evidence shows that student has some interactions with others related to the academic skills and/or in work, social, and/or community situations.	Little or no evidence that student has interactions with others related to academic skills and/or in work, social, and/or community situations.	

APPENDIX C. WAAS PORTFOLIO ACADEMIC ACHIEVEMENT STANDARD DESCRIPTIONS

The academic achievement standards for students with significant disabilities who are participating in the Washington Alternate Assessment System (WAAS) portfolio are significantly different from the standards for students who participate in the Washington Assessment of Student Learning (WASL). The WAAS portfolio is based on the Essential Academic Learning Requirements (EALR) Extensions, which allow the student to participate and progress in the general curriculum. Because the WAAS portfolio is based on the student’s Individualized Education Program (IEP) goals in relation to the EALR Extensions, the specific assessment targets selected for the student may be the same for many content areas but may be different for any other student. Additionally, these students have educational goals that may remain the same throughout their educational careers. Therefore, the following academic achievement standard descriptors apply for all grades and content areas.

<p>Level 4: Exceeds Standard</p>	<p>Evidence in the portfolio shows that the student is meeting or exceeding her/his goals for both academic skills. The evidence also shows that the student applies the academic skills in three or more contexts; therefore, achievement shown in the portfolio is a reliable (dependable) demonstration of the skills linked to the targeted GLEs.</p>
<p>Level 3: Meets Standard</p>	<p>Evidence in the portfolio shows that the student meets or exceeds her/his goal for one academic skill and approaches her/his goal for the second academic skill. The evidence also shows that the student applies both academic skills in two or more contexts; therefore, achievement shown in the portfolio is a mostly reliable demonstration of the skills linked to the targeted GLEs.</p>
<p>Level 2: Approaches Standard</p>	<p>Evidence in the portfolio shows that the student is approaching her/his goals for both academic skills. The evidence also shows that the student applies the academic skills in only one or two contexts; therefore, achievement shown in the portfolio is only somewhat reliable as a demonstration of the skills linked to the targeted GLEs.</p>
<p>Level 3: Well Below Standard</p>	<p>Evidence in the portfolio suggests that the student is nowhere near meeting her/his goals for the academic skills. The evidence shows that the student applies each academic skill in only one context; therefore, any achievement shown in the portfolio is not a reliable demonstration of the skills linked to the targeted GLEs.</p>