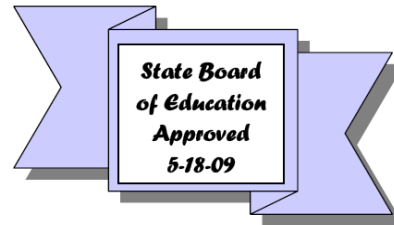# SUMMARY REPORT

ACHIEVEMENT LEVELS
FOR
ARIZONA'S INSTRUMENT to MEASURE STANDARDS ALTERNATE (AIMS A)

For
Arizona Department of Education
Exceptional Student Services

State Board
of Education
Approved
5-18-09

May 30, 2009

From May 14 to May 16, 2009, a Standard Setting Session was held in Phoenix with 37 Arizona educators to: (1) establish achievement levels for students with disabilities, in Grades 3 through 8 and 10, who participated in the Arizona Alternate Assessment (Arizona's Instrument to Measure Standards Alternate, AIMS A) and (2) refine the performance level descriptors for each grade level and content area assessed. The session was led by Stephen N. Elliott from Vanderbilt University with assistance from Arizona Department of Education (ADE) personnel Roberta Alley, Dr. Leila Williams, Danielle Gordon, Melanie Mosiman, Dr. Charles Bruen, Marilee Beach, and Forster Okoli. A copy of the agenda for this meeting is provided as ***Appendix A***. The results from this Standard Setting Session are summarized in this document and are offered as recommendations to guide Arizona educational leaders' decisions for determining achievement levels on AIMS A in Reading, Mathematics, and Science for over 6,400 students with significant disabilities.

Overview of Standard Setting

Standard Setting is the process of determining appropriate achievement levels that correspond to a specified level of proficiency. The purpose is to establish achievement levels that are based on what students in each achievement level should know and be able to perform. For example, if a student obtained or exceeded the achievement level corresponding to the "Meets" level, then that student should have demonstrated knowledge, skills, and competencies sufficient to be called "proficient" for AYP purposes. This requires the participant to first specify what a proficient student should be expected to understand and perform, and then to determine the achievement levels that correspond to those expectations.

Besides deriving achievement levels for each content area, this process yields descriptions of what students who achieve the various achievement levels typically know and are able to perform. By examining the description of students' typical performances in a given achievement level, one gains an understanding of the knowledge, skills, and abilities typically held by students in that level and identify skills that a given student is not yet able to perform consistently. This type of information helps teachers communicate with others about a student's progress, next year's instructional goals, and the status of the student relative to the state's learning standards.

There is a good deal of judgment involved in Standard Setting and a need to establish a high level of confidence in these judgments. Thus, it is important to have a representative group of educators familiar with the curricular and instructional needs of students with significant disabilities and also knowledgeable of the current alternate assessment to participate on a Standard Setting Panel. It is also typical to have several general educators knowledgeable of the state's academic standards and curriculum, and a few parents of students with significant disabilities on the committee.

AIMS A includes Reading, Mathematics, and Science tests. At each grade level, 3-8 and 10, there are 20 Reading items and 22 Mathematics items, respectively. The Science test, which is administered at grades 4, 8, and 10, also has 20 items. Each item on each test at every grade level is worth 4 points. Thus, scores on the Reading and Science tests range from 0 to 80, while scores on the Mathematics test range from 0 to 88. The primary objective of the Standard Setting Panel was to determine where along the score

continuums in each content area, the score or cut point would be for a marginally proficient student. In other words, the panel's main job was to determine "how many score points was enough" to be deemed to "meet the standard" in reading, mathematics and science in each tested grade. Once the "Meets" cut point was established for a grade level test, the panel determined the cut points for the "Approaches" and "Exceeds" achievement levels at that same grade level.

The Bookmark Procedure

Several different approaches to establishing achievement standards exist. An item mapping method referred to as the Bookmark Procedure was utilized to establish the achievement (performance) standards for AIMS A for students with significant disabilities. The Bookmark Procedure (Lewis, Mitzel, & Green, 1996) was developed by researchers at CTB/McGraw-Hill and has been used to establish the achievement standards for many states' regular achievement tests and several states' alternate assessments over the past decade. This procedure is recognized as a scientifically defensible procedure by the USDE. Standard Setting using this procedure involves presenting experienced educators a booklet with a set of test items ordered from easiest to most difficult. A separate test booklet of items is presented for each content area (i.e., reading, mathematics, and science) and an item map with item difficulty data accompanies the test item booklet. After carefully studying the ordered-items in a booklet, a unique achievement level for a given achievement (performance) level is identified. The participants determine the achievement level by placing a bookmark at the location in the booklet where they think a student who is functioning at a given level will likely respond successfully to items preceding the bookmark. Items preceding the bookmark represent content that all proficient students should likely know and perform. The final achievement level is computed as the median of the number of items immediately before and after the bookmark. Although this sounds quite simple, in fact, committee members often expend considerable effort in reaching their final decisions about the knowledge, skills, and competencies needed to be considered "proficient."

A general description of the steps involved in the Bookmark Procedure for each of the content areas in AIMS A follows:
- ❑ Introduction to Standard Setting
- ❑ Review all Items on the assessment
- ❑ Review and discuss the current Performance Level Descriptors for each achievement level
- ❑ Reach Consensus on the definition of "Meets the Standard" as measured by AIMS A
- ❑ Round 1: Individuals independently place marks in test booklets to indicate "Meets the Standard" achievement level
- ❑ Post-Round 1: Individuals at each table discuss their placements of marks for the "Meets the Standard" achievement level
- ❑ Round 2: Teams at each table make a consensus decision about marks for the "Meets the Standard" achievement level
- ❑ Post-Round 2: Feedback is provided about the median achievement levels and the likely distribution of students at each level, then the group can discuss rationale for their ratings
- ❑ Round 3: Teams collectively make final decisions about marks for each of four levels of Achievement
- ❑ Post-Round 3: Feedback is provided about the Committee's Median Achievement levels and likely impact on student distributions
- ❑ Review and discuss the trends across grade levels for a given content area and examine any significant outliers

❑   Review and revise, if necessary, the descriptions associated with each of the four levels of achievement

The three-round Bookmarking procedure was followed for each content area assessed by AIMS A for Grades 3, 4, 5, 6, 7, 8, and 10 in Reading, Mathematics, and Science. The outcome of this Bookmark procedure resulted in identified achievement levels for each of the grade-level content areas on AIMS A. The detailed result of what constitutes a "proficient performance" on AIMS A contributes information that can be integrated with other students' results on AIMS to be used for school accountability. Together the results from AIMS A and AIMS provide assessment data for all students in Arizona Public Education Associations (PEAs) for the federally required adequate yearly progress (AYP) calculation and report.

Participants and Group Assignments

The 37 participants in the Standard Setting Session represented educators from school districts and educational agencies from across the state. All the participants were familiar with, or had experience administering, AIMS A. The participants and their professional affiliations are listed in **Appendix B**. These participants formed nine teams representing elementary, middle and high schools who worked together for the entire 3-day session. Five teams had four members (three special educators and a regular educator or dual certified educator) and four teams had five members (four special educators and a regular educator or dual certified educator). This team structure was designed to enhance the developmental sensitivity and representativeness of the team's decisions. Three groups of participants – elementary, middle school, and high school – were created to determine cut scores for each grade and content area. To improve consistency and achieve equity in the recommended cut scores across the 3-8 and 10 grade-spans, a cross-lag design with different groups of teachers was used to ensure independent replications of Reading and Mathematics cut scores in grades 5 and 7, and for Science in all grades 4, 8, and 10. A visual of the three groups and their various grade and content assignments is provided as Figure 1.

Figure 1. Overview of the AIMS A Standard Setting Session and Grade-Level Teams
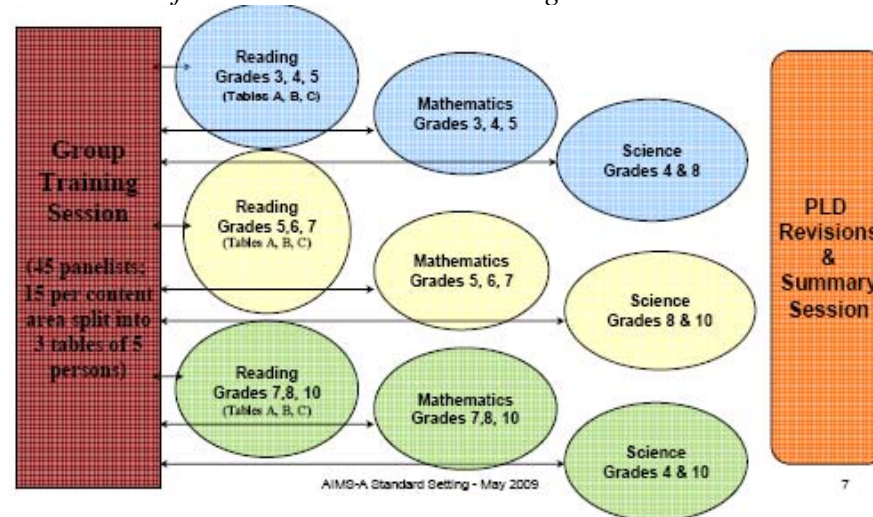
Overview of the Students of Interest

The sample of students in the AIMS A database at each grade level averaged 870 per grade with a range from 798 (6th grade) to 1368 (high school) students in prescribed assessment years and is representative of the state's school age population. Students eligible to take AIMS A were all identified with approved criteria that included having a significant disability and functioning several grade levels below their age mate peers with milder disabilities. The majority of the students qualifying to take AIMS A has been receiving special education services since entering school and has been classified as moderate or severely mentally retarded, or autistic. These students have been receiving instruction based on the Arizona Alternate Academic Standards and have been determined to need significantly more accommodations than allowed to take AIMS.

Definition of Proficient (Meets the Standard)

One of the most important steps in Standard Setting is to achieve a consensus definition of what it means to be "proficient." Once a consensus definition of proficient is determined, it provides a foundation for making decisions about the knowledge and skills that a student should be able to demonstrate if they are to be considered proficient. The participants in the Standard Setting Session spent considerable time discussing what it means for a student to be *proficient* or in the terms of the Arizona Achievement Standards to *meet the standard*. To facilitate their thinking about this definition, they were provided the performance level descriptors approved by the Arizona State Board of Education, a copy of the state's content standards for students with significant disabilities, and a copy of AIMS A items. The state's four achievement levels for each of the content areas assessed by AIMS A are documented in *Appendix C*. These achievement levels were a centerpiece of the Standard Setting Training Session (see *Appendix D* for training slides).

Materials and Decisions about Achievement levels

The key materials used to conduct the Standard Setting were ordered item test booklets, item maps with AIMS A items from each content area rank ordered by difficulty from easiest to hardest (see *Appendix E*), and item graphs (see *Appendix F*) and item tables (see *Appendix G*) portraying the total score distributions of students who were administered AIMS A in spring 2009. An example of the item map for AIMS A Reading is displayed in *Appendix E*. Figures 2, 3, and 4 provide score distributions for the 4th, 8th, and 10th grade AIMS A Reading test. These distributions are illustrative of those in Mathematics and Science at the same grades and indicate AIMS A overall is a difficult test for about 15% of eligible students. Some students, however, also do very well on the tests.
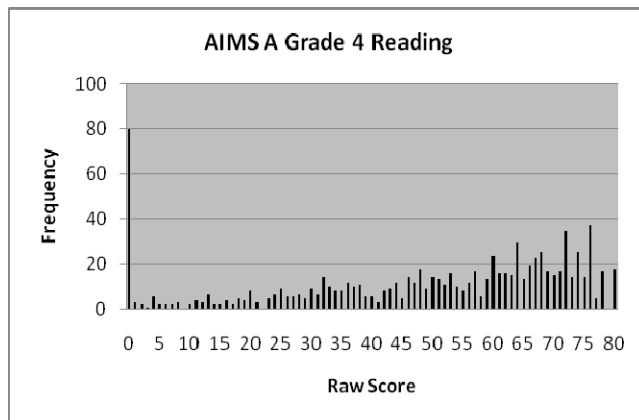
Figure 2.Grade 4 Reading   *Figure 3. Grade 8 Reading*
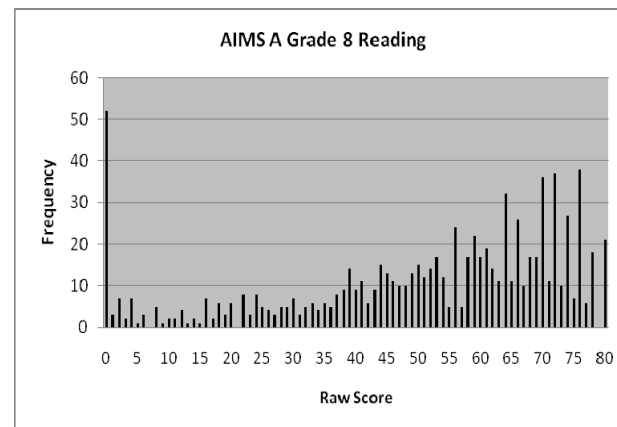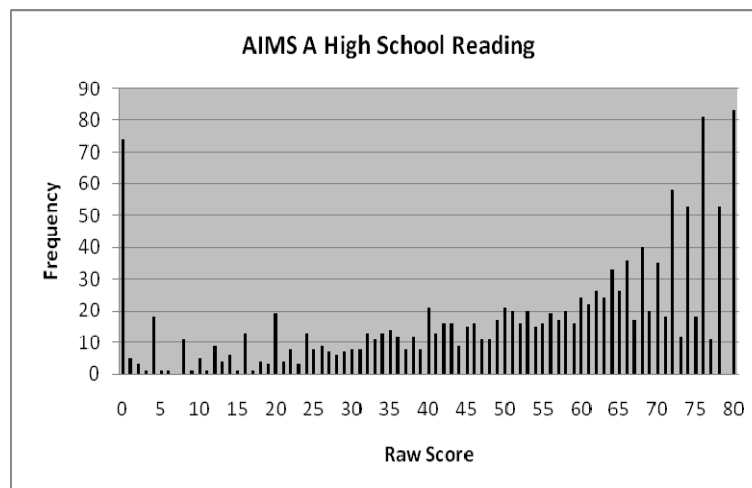

Figure 4. High School Reading

To facilitate communication and decision-making about AIMS A Standard Setting outcomes, the following assumptions were stated and agreed upon by all participants at the outset of the process:

- Arizona's academic achievement levels are Falls Far Below, Approaches, Meets, and Exceeds the Standard.

- The 4 levels of achievement for a given content area need <u>not</u> be equal in nature; that is, they need not cover the same number of items or possible points, nor do they necessarily need to represent an equal proportion of students.

- There are likely developmental differences that should be considered when setting performance standards.

- Given the need to yield overall decisions of "proficient" or "not proficient yet" for AYP, a single number for a achievement level must be determined even though we know that all scores have some error and it is best professional practice to provide a confidence band around a score. To off-set concerns about error in a single score, it is recommended that important decisions be based on more than one test score.

- Different people reviewing the same items and same impact data might reasonably derive somewhat different achievement levels in the three content areas. Therefore, to establish confidence in the recommended cut scores a replication method was employed at a subset of grades (i.e., 5 and 7 for Reading and Math, and 4, 8, and 10 for Science) whereby "second" teams of educators independently set cut scores.

- The results of the Standard Setting Process would be presented to the Arizona State Board of Education as recommendations to follow when determining whether or not a student meets the standard (e.g., is proficient) on AIMS A. Thus, the participants' recommendations are advisory.

After reading the consensus definition of *meets the standard*, participants used the rank-ordered item tables to record their decisions about what alternate knowledge and skills it took to be considered *proficient*. Participants first made independent decisions about the number of items it would take to *meet the standard*, then worked with their tablemates to reach a consensus on the number of items that it would take to *meet the standard*. Once all the table leaders reported a consensus number of items for the *meets the standards* level, the median number of items needed to *meets the standard* as defined by all tables was determined. Once this achievement level was determined, it served as the "Meets the Standard" achievement level for the content area, and then impact data were provided via the cumulative score distribution figures. To operationalize impact, all participants were provided a cumulative frequency distribution with the percentage of students likely to be considered as *meeting the standard* in a content area. In some cases, participants requested comparison data for students on AIMS. The consensus achievement level and impact data collectively were discussed among the entire group of participants and a final decision was made about an achievement level at each grade level for a given content area.

After reaching a final decision about the *meets the standard* achievement level for each area, teams were asked to determine the achievement levels differentiating AIMS A performances at the Falls Far Below level from the Approaches level, and the Meets level

from the Exceeds level of achievement. For these decisions, an abbreviated version of Bookmarking featuring only the table consensus decisions with impact data as feedback was used to determine median cut points.

Finally, after all cut point recommendations for each content area in each grade were completed, an integrated review of the suggested cut points and related impact data across all grades was presented to the participants by the session leader. This review focused on consistency across grades for a given content area. Given that the numbers of possible score points were the same across grades within content areas, it was easy to identify outliers by looking at both the recommended cut scores and the likely percentage of students "passing" rates. Using this approach, the cut score for the *meets the standard* level for Reading at grades 3 and 10 were considered relatively low and the cut score for Mathematics at grade 5 was considered relatively high. The respective teams that set the original cut scores agreed to review their recommendations. The outcomes of these reviews were adjustments that resulted in cut scores that were more consistent with those for the same content area at other grades.

Standard Setting Results

The results of the 2009 AIMS A Standard Setting Session are summarized in a series of tables (1, 2, and 3) and figures (5 through 10) that follow. The initial table for each content area provides the recommended raw cut scores for at each grade level for the four achievement levels. These tables also provide impact data in the form of the number and percentage of students that would be at each achievement level in each grade in 2009, if these cut scores were adopted. The accompanying figures simply provide a visual depiction of the same data for each content area. Finally, an integrated summary table is provided of the raw score ranges for each achievement levels in a given content area. Please note that AIMS A tests have different items and different performance level descriptors (PLDs) for each grade level.

Table 1. AIMS A Reading Recommended Cut Score								
Grade	**3**	**4**	**5**	**6**	**7**	**8**		**10**
Far Below	0 - 20	0 - 16	0 - 12	0 - 12	0 - 15	0 - 16		0 - 12
Approaches	21 - 40	17 - 44	13 - 42	13 - 40	16 - 39	17 - 40		13 - 40
Meets	41 - 64	45 - 70	43 - 68	41 - 66	40 - 67	41 - 70		41 - 72
Exceeds	65 - 80	71- 80	69 - 80	67 - 80	68 - 80	71 - 80		73 - 80
Number of Students								
Grade	**3**	**4**	**5**	**6**	**7**	**8**		**10**
Far Below	141	125	89	104	112	100		130
Approaches	156	200	209	193	141	134		244
Meets	384	391	334	335	346	451		683
Exceeds	196	182	175	166	205	175		311
Total	877	898	807	798	804	860		1368
Percentage of Students								
Grade	**3**	**4**	**5**	**6**	**7**	**8**		**10**
Far Below	16.08	13.9	11.02	13.04	13.91	11.63		9.5
Approaches	17.78	22.27	23.92	24.2	17.55	15.6		17.8
Meets	43.76	43.53	43.37	41.96	43.04	52.46		49.91
Exceeds	22.33	20.26	21.69	20.8	25.5	20.34		22.73

Figure 5. Reading Recommended Cut Scores Across the Grades
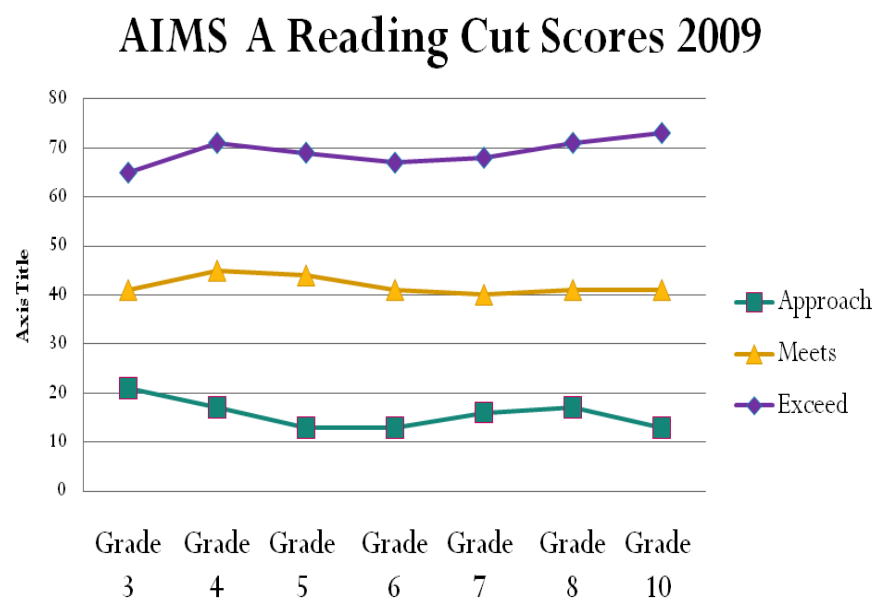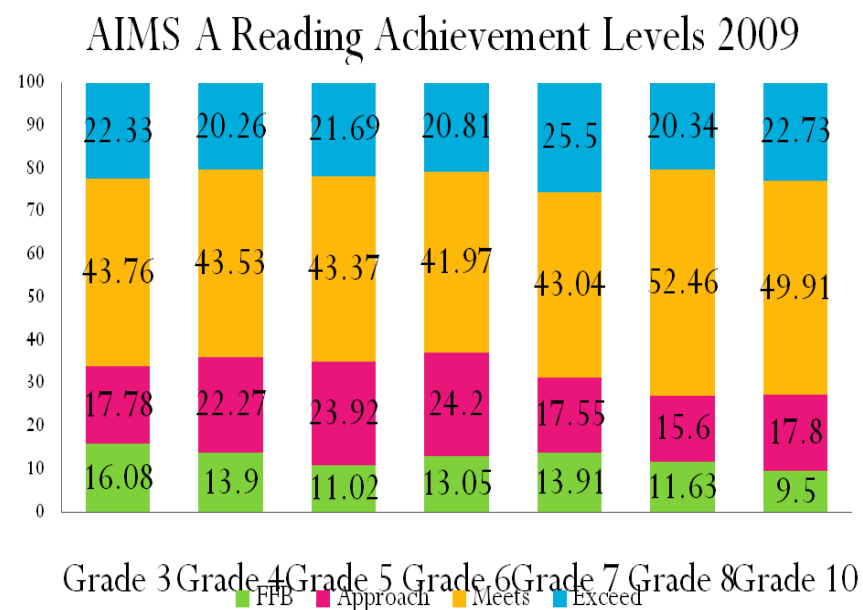
Figure 6. Percentage of Students at Each Reading Achievement Level

Table 2. AIMS A Mathematics Recommended Cut Score								
Grade	**3**	**4**	**5**	**6**	**7**	**8**		**10**
Far Below	0 - 20	0 - 20	0 - 20	0 - 16	0 - 12	0 - 16		0 - 16
Approaches	21 - 40	21 - 40	21 - 40	17 - 44	13 - 40	17 - 40		17 - 40
Meets	41 - 72	41 - 72	41 - 72	45 - 72	41 - 72	41 - 68		41 - 76
Exceeds	73 - 88	73 - 88	73 - 88	73 - 88	73 - 88	69 - 88		77 - 88
Number of Students								
Grade	**3**	**4**	**5**	**6**	**7**	**8**		**10**
Far Below	130	168	146	106	92	115		192
Approaches	142	140	149	202	166	185		293
Meets	388	358	399	366	387	360		728
Exceeds	217	232	113	124	159	200		155
Total	877	898	807	798	804	860		1368
Percentage of Students								
Grade	**3**	**4**	**5**	**6**	**7**	**8**		**10**
Far Below	14.8	18.7	18.08	13.33	11.44	13.4		14.03
Approaches	16.17	15.59	18.48	25.34	20.64	21.51		21.42
Meets	44.22	39.85	49.46	45.89	48.15	41.85		53.23
Exceeds	24.74	25.82	14.00	15.55	19.79	23.26		11.33

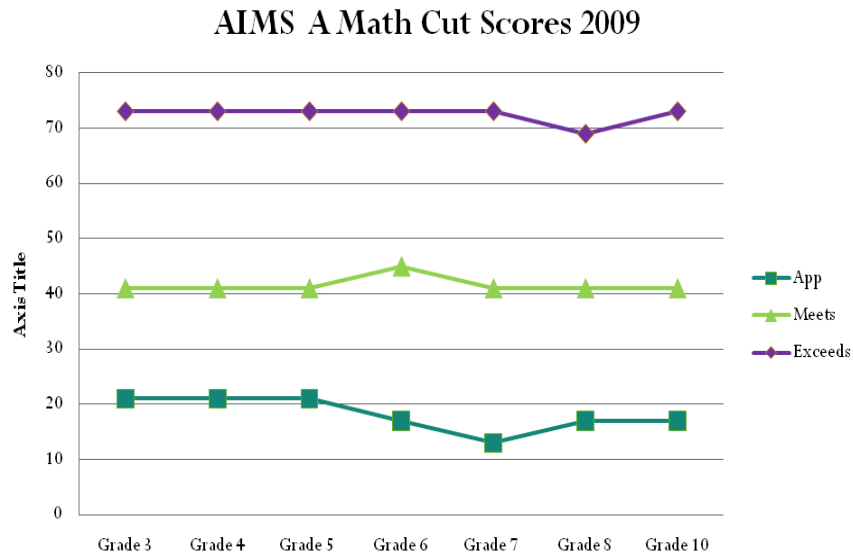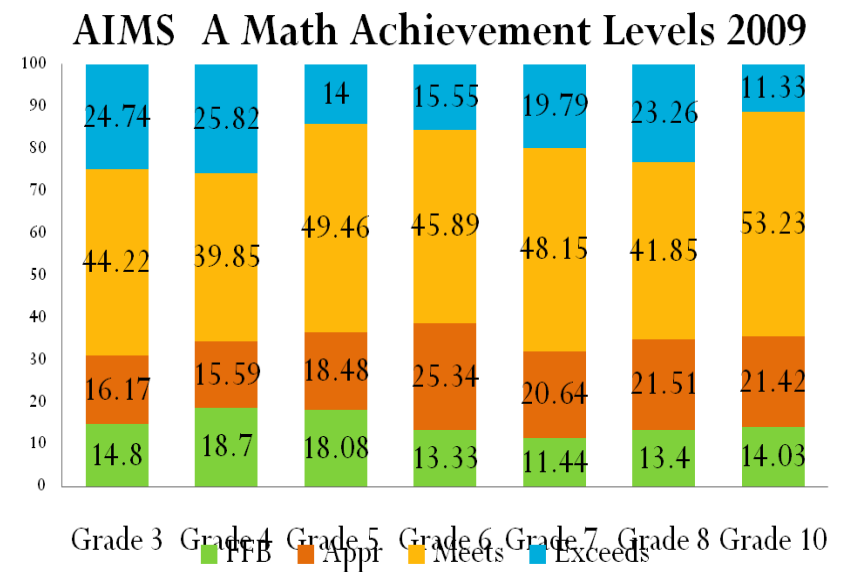*Figure 7. Mathematics Recommended Cut Scores Across the Grades*

Figure 8. Percentage of Students at Each Mathematics Achievement Level

| | | | | Table 3. AIMS A Science Recommended Cut Score | | | | |
Grade	3	4	5	6	7	8		10
Far Below		0 - 14				0 - 16		0 - 12
Approaches		15 - 44				17 - 45		13 - 42
Meets		45 - 72				46 - 74		43 - 70
Exceeds		73 - 80				75 - 80		71 - 80
				Number of Students				
Grade	3	4	5	6	7	8		10
Far Below		119				85		80
Approaches		181				141		156
Meets		388				393		378
Exceeds		209				241		207
Total		897				860		821
				Percentage of Students				
Grade	3	4	5	6	7	8		10
Far Below		13.24				9.91		9.73
Approaches		20.15				16.41		19.01
Meets		43.25				45.72		46.04
Exceeds		23.3				28.01		25.2

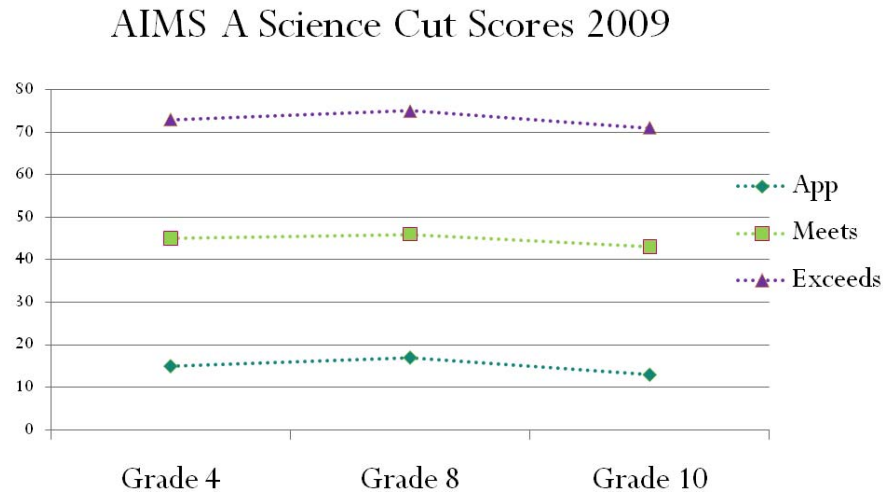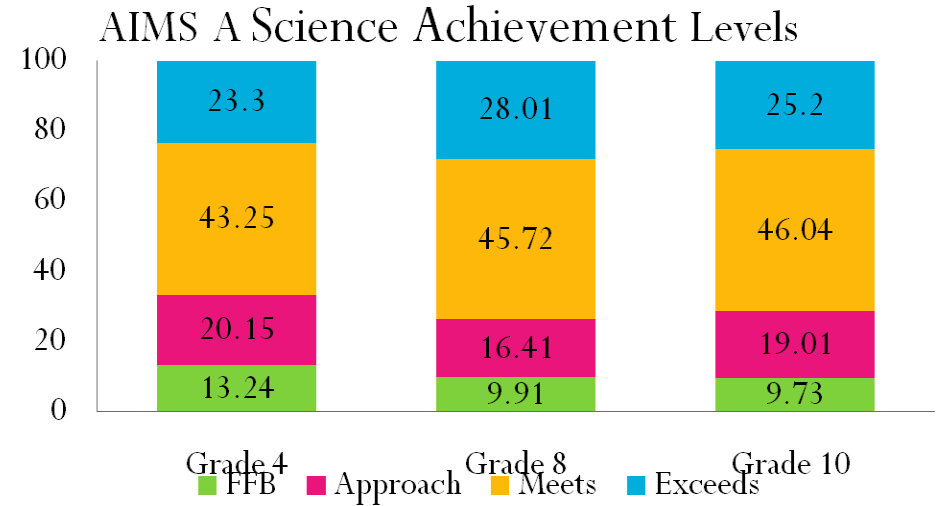*Figure 9. Science Recommended Cut Scores Across the Grades*

Figure 10. Percentage of Students at Each Science Achievement Level

AIMS A Science Cut Scores 2009

AIMS A Science Achievement Levels

The following principles guided the development of final cut scores for AIMS A achievement levels for each content area:

➢ Creditable assessment systems for interpreting student achievement should reflect general developmental trends and instructional expectations whereby older or more advanced students, on average, consistently exhibit more knowledge and skills in a given content area. Given the design of AIMS A where there are an equal number of items on each test and these items are based on grade-sequenced extended content standards, it was expected that cut scores across grades for the same content would be very similar.

➢ The recommended cut scores can be conceptualized with a confidence band of $\pm$ 5 raw score points based on what is known about the standard error of measurement for the tests. Given it is an accepted scientific practice to use confidence or error bands around scores when making important decisions, the panel supported the application of such a band for the purposes of making final adjustments to cut scores. However, this adjustment procedure was not necessary because the recommended cut scores were quite uniform and conformed to the expected developmental trends.

The recommended achievement levels for AIMS A Reading, Mathematics, and Science followed these guidelines and are intended to be of use to educators, parents, and other educational stakeholders interested in the achievement of students with significant disabilities. At the conclusion of the Standard Setting Session, the data featured in Tables 1, 2, and 3 were presented and discussed among all three grade-level groups of panelists. The result was that <u>panel members unanimously endorsed the cut scores</u> documented in this report.

Following the endorsement of the cut scores, panelists revisited the Performance Level Descriptors for each content area and grade level with the purpose of documenting ways to improve them as communication tools. After the Arizona State Board of Education approval of the final cut scores, further refinements to the PLDs become possible by using the item maps to identify discriminating items just beyond cut scores. These items can then be added to the PLDs to provide a comprehensive description of what it means to *meet the standard* for students with significant cognitive disabilities in Arizona.

Reference

Lewis, D.M., Mitzel, H.C., & Green, D.R. (1996, June). Standard setting: A Bookmark approach. In D.R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring*. Symposium conducted at the meeting of the Council of Chief State School Officers National Conference on Large Scale Assessment, Phoenix, AZ.

About the Primary Author of this Report

Stephen N. Elliott received his doctorate at Arizona State University in 1980 and is a Professor of Special Education and the Dunn Family Chair of Educational and Psychological Assessment in Peabody College at Vanderbilt University. Steve teaches courses on the measurement and assessment of academic and social behavior. He currently co-directs three USDE research grants concerning the validity of testing modifications and alternate assessments for students with disabilities. He also directs Peabody College's Interdisciplinary Program in Educational Psychology and serves as the Director of the Learning Sciences Institute, a trans-institutional center for externally funded research. He has authored more than 140 journal articles, 20 books, 35 chapters, and 5 widely used behavior-rating scales. His research focuses on scale development and (a) the assessment of children's social skills and academic competence and (b) the use of testing accommodations and alternate assessment methods for evaluating the academic performance of students with disabilities for purposes of educational accountability. Steve has helped design alternate assessments in several states (HI, ID, MS, & WI) and has led Standard Settings in each of these states for these assessments of students with significant disabilities. In 2009, he was named a Fellow in the American Educational Research Association and selected as a Senior Scientist for Division 16 of the American Psychological Association.

Appendix A

Session Agenda

AIMS A 2009 Standard Setting

<p style="text-align:center">**Standard Setting Workshop**
Arizona Alternate Assessment – AIMS A
May 14 - 16, 2009</p>

Leaders: Stephen Elliott, Vanderbilt University   **Location**: Sheraton Crescent
                Roberta Alley, ADE   2620 W. Dunlap Avenue
                Charles Bruen, ADE   Phoenix, AZ
                Danielle Gordon, ADE
                Leila Williams, ADE

Thursday, May 14

8:30 a.m.   **Welcome/Introductions**

8:45 a.m.   **Non-Disclosure and Travel Procedures**

9:00 a.m. – 12:00 p.m. **Workshop**

- ❑ **Workshop Goals and Role of Participants**
    Goal #1   *Review the AIMS A items and the related statistics for science,*
                          reading, and mathematics items for grades 3 through 8 and 10 and
                          impact data based on 2009 results.

    Goal #2   *Establish recommended proficiency cut-scores for AIMS A science,*
                          reading, and mathematics assessments for students with significant
                          disabilities in grades 3 through 8 and 10.

- ❑ **Background of Arizona's Statewide Assessment & Accountability System**

- ❑ **Introduction to Standard Setting: Rationale and the Bookmarking Procedure**
    - o Activity: Connecting PLDs to Item Maps
    - o Defining the Marginally Proficient Student

- ❑ **Major Steps in a Modified Bookmark Procedure**

Review and complete all AIMS-A Multiple Choice and Rating Scale Items
Review and Discuss current Performance Level Descriptors for each achievement level
Reach Consensus on the definition of "Meets the Standard"
Round 1: Individual Proficiency Cut-Point Determination
Post-Round #1 Discussion
Round 2: Team Consensus for Proficiency Cut-Point
Post-Round #2 Discussions with Feedback on Impact
Round 3: Teams Final Decisions
Post Round #3: Feedback & likely impact on student distributions
Review and Revise Proficiency Level Descriptors
Committee Recommendations to the State Board of Education for approval and adoption

❑ **Table Assignments & Decision Making Guidelines**

Thursday, May 14 (1:00 a.m. – 5:00 p.m.)

❑ **Review Standard Setting Procedures and Discuss Issues**

❑ Review the AIMS-A Reading Items grades 3, 4 & 5 and Conduct Standard Setting

❑ Review the AIMS-A Reading Items grades 5, 6 & 7 and Conduct Standard Setting

❑ Review the AIMS-A Reading Items for 7, 8, & High School and Conduct Standard Setting

Friday, May 15 (8:00 a.m. – 5:00 p.m.)

❑ **Review Standard Setting Procedures and Discuss Issues**

❑ **Complete Review of AIMS A Reading Items**

❑ Review the AIMS-A Mathematics Items grades 3, 4 & 5 and Conduct Standard Setting

- ❑ Review the AIMS-A Mathematics Items grades 5, 6 & 7 and Conduct Standard Setting

- ❑ Review the AIMS-A Mathematics Items for grades 7, 8, & High School and Conduct Standard Setting

Saturday, May 16 (8:00 a.m. – 5:00 p.m)

- ❑ **Review Standard Setting Procedures and Discuss Issues**

- ❑ **Complete Review of AIMS A Mathematics Items**

- ❑ Review the AIMS-A Science grade 4 & 8 and Conduct Standard Setting

- ❑ Review the AIMS-A Science grades 8 & 10 Conduct Standard Setting

- ❑ Review the AIMS-A Science for grades10 & 4 Conduct Standard Setting

- ❑ **Suggestions for Refining AIMS-A Performance Level Descriptors**

- ❑ **Review Results of Standard Setting Workshop**

- ❑ **Participant Evaluations**

Appendix B

Participants in the 2009 Standard Setting for AIMS A

Last	First	Race	Sex	Title or Occupation	Certification	District
Adams-Brown	Susan	B	F	Resource Teacher Middle School; 7 & 8 Social Studies; Language Arts, Math	Elementary/Spec Cross Cat	Cartwright Elementary District
Andersen	Tamara	B	F	Special Education K-5, Self Contained	Special Education MIMR K-12	Tolleson Elementary District
Apuna	Sandra	W	F	District Language Arts Coordinator	Elementary/Junior High School / Special Education	Gilbert Unified District
Barsevich	Valerie	W	F	Sixth Grade - Mathematics	Elem & Spec Ed Mentally Hand Certif./Principalship	Tucson Unified School District
Bates	Heather	W	F	Freshman English Teacher and Junior English Teacher	Secondary, English and Special Education,CCS	Tucson Unified School District
Bonney-Clay	Mepet	W	F	High School Self Contained Spec Education Teacher age (14-21)	Cross Categorical Special Education	Parker Unified School District
Cassidy	Kay	A	F	Retired	Secondary	(blank)
Cox	Rebecca	W	F	Primary Special Education/ Supervision of RTI Program Grades K-3	Elementary/Special Education	Flagstaff Unified District
Csurka	Lucy	W	F	Jr High Art and Reading /7th Grade Reading	Secondary 7-12; Art K12, Spec Educ K12, LD/MR/SelfContained	Theodore Roosevelt School
D'Antonio-Schleich	Peggy	W	F	Special Education Teacher	Special education-Cross Categorical	Phoenix Union High School District
Dumas	Donna	W	F	Retired	BS Elementary, K-8th, Special Educ., MA Administration	(blank)
Duncan	Elizabeth	W	F	Int. MOMR, Self Contained Teacher	Special Ed, Elementary	Roosevelt Elementary District
Faiveley	Patricia	W	F	4th Grade all subjects	Elementary, Special Education.	Scottsdale Unified District
Fetter	Kathy	W	F	Spec Educ Cross Categorical Spec Class K-2 Teacher__	Standard Spec Educ LD K-12; Stand Spec Educ MR K-12; Provisional Struct English Imm Endorsement K-12	Amphitheater Unified District
Fortier	Jacqueline	H	F	Teacher of Moderately Cognitively Impaired 9-12	Secondary Certification, Special education.	Tucson Unified School District

Franklin	Rebecca	W	F	Teacher 9-12+ Grade Self-Contained Life Skills Prgm, MIMR,MOMR,	Special Education - Arizona	Kingman Unified School District
Fritsche	Janice	W	F	High School Special Services	Cross Cat K-12, severely profound k-12, ,OTR	Douglas Unified School District
Geiger	Vicki	W	F	Education Prgm Specialist- Special Education @ State Hospital & Adult Educational services through Rio Salado	Reg Education K-8, Special Educ K-12 ED and LD	Arizona State Hospital
Hammond	Mary Jo	W	F	K-5 Language Arts resource room	Elementary/Special Education	Kingman Unified School District
Hart	Holly	W	F	5/6th Grade Cross Categorical Self Contained	Special Education	Washington Elementary District
Hebein	Jenna	W	F	self contained 3rd grade cross categorical developmental class (MIMR-MOMR)	Elementary, cross cat. Special Education, severe/profound special ed	Washington Elementary District
Hellerud	Linda	W	F	H.S. Special Education -MIMR, Resource Room	Spec Ed, Mental Retardation, Learning Disabilities	Colorado River Union High School District
Johnson	Jennifer	W	F	Special Education Facilitator	Elem, Secondary, Sp Ed: Cross-Categorical K-12, Severe & Profound Disabilities, English, History	Amphitheater Unified District
Morrow	Karin	W	F	Self-Contained MI/MO High School	Cross-Cat Sped K-12, Elem. Ed. K-8	Dysart Unified District
Mosiman	Michael	W	M	Resource and Self-Contained ED/MIMR	Special Education K-12	Tempe Elementary District
Peaslee	Kimberly	W	F	High School Instructional Specialist 9-12	Special Education / Principal	Phoenix Union High School District
Pyle	David	W	M	Teacher, Self-contained 5-8, Reading, Math and Written Expression	Special Education K-12, Principal	
Roth	Natalie	W	F	Reading and Math; Gifted 3-6 Teacher	K-12; Drama & Speech; Principal	Deer Valley Unified District
Sholl	Shyla	H	F	Self-Contained, Cross-Categorical Special Education Teacher 3-5	Elementary K-8 and Spec Education, Cross-Categorical K-12	Amphitheater Unified District
Sims	Kimberly	H	F	Working on doctoral studies Educational Leadership & Teacher Innovation	Spec Education K-12, LD, ED, MR	Student-doctoral degree

Stair	Carin	W	F	K-5 resource teacher	Stand SpEd Learning Disabilities k-12; Mental Retardation; Administrative Certificate/Principal SEI	Tucson Unified School District
Swartz	Najah	NA	F	Hearing Impaired Itinerant Teacher K-12	Hearing Impaired k-12 Special Education	Tucson Unified School District
Thompson	Loriann	W	F	H.S. Severe Autism Program	SpEd ED, LD, OHI, SMR, MR	Tempe Union High School District
Tiernan	Maureen	W	F	9th – 12th grade Medical Fragile	K-12 Special Education	Phoenix Union High School District
Walch	Betty	W	F	Retired	Special Ed. Secondary, Administrative.	(blank)
Whitaker	Johanna	B	F	3-7 cross-categorical moderate-severe/behavioral	Cross Categorical K-12	Washington Elementary District
Williams	Christina	W	F	Inclusion Specialist	Spec. Ed. K-12, Severe/ Profound	Vail Unified School District

Appendix C

Example Performance Level Descriptors

Arizona Alternate Standard Performance Level Descriptors
Grade 4 Reading

Exceeds the Standard – Students with significant cognitive disabilities who score in this level can typically function independently or with minimal cueing to demonstrate mastery of subject matter as reflected by the alternate reading standard.

Meets the Standard – Students with significant cognitive disabilities who score in this level can typically function with moderate support through the use of visual representations, manipulatives, and objects to demonstrate a solid understanding of subject matter as reflected by the alternate reading standard.

Approaches the Standard – Students with significant cognitive disabilities who score in this level can typically function with extensive support through the use of visual representations, manipulatives, and objects to demonstrate partial understanding of subject matter as reflected by the alternate reading standard.

Falls Far Below the Standard – Students with significant cognitive disabilities who score in this level may have significant gaps and limited knowledge and skills that are necessary to satisfactorily meet the state's alternate reading standard. Students will typically require a considerable amount of additional instruction and intervention in order to achieve a satisfactory level of understanding.

Students at the "**Exceeds the Standard**" level generally know the skills required at the "Meets" and "Approaches" levels and are able to:	Students at the "**Meets the Standard**" level generally know the skills required at the "Approaches" level and are able to:	Students at the "**Approaches the Standard**" level generally know and are able to:
• Follow a set of multi-step directions in order. • Identify specific facts in text. • Select a synonym, antonym, and homonym. • Make a prediction.	• Locate information from functional text. • Determine meaning of a simple or environmental word. • Identify the conflict or problem.	• Identify cause and effect. • Find a solution to a problem. • Identify one aspect of the setting. • Describe a character's trait.
80-- 71	70-- 45	44-- 17

These descriptors do not include all the skills and knowledge as contained in the Alternate Reading Standard.

Arizona Alternate Standard Performance Level Descriptors
Grade 4 Mathematics

Exceeds the Standard – Students with significant cognitive disabilities who score in this level can typically function independently or with minimal cueing to demonstrate mastery of subject matter as reflected by the alternate mathematics standard.

Meets the Standard – Students with significant cognitive disabilities who score in this level can typically function with moderate support through the use of visual representations, manipulatives, and calculators to demonstrate a solid understanding of subject matter as reflected by the alternate mathematics standard.

Approaches the Standard – Students with significant cognitive disabilities who score in this level can typically function with extensive support through the use of visual representations, manipulatives, and calculators to demonstrate partial understanding of subject matter as reflected by the alternate mathematics standard.

Falls Far Below the Standard – Students with significant cognitive disabilities who score in this level may have significant gaps and limited knowledge and skills that are necessary to satisfactorily meet the state's alternate mathematics standard. Students will typically require a considerable amount of additional instruction and intervention in order to achieve a satisfactory level of understanding.

Students at the "**Exceeds the Standard**" level generally know the skills required at the "Meets" and "Approaches" levels and are able to:	Students at the "**Meets the Standard**" level generally know the skills required at the "Approaches" level and are able to:	Students at the "**Approaches the Standard**" level generally know and are able to:
• Subtract whole numbers. • Add whole numbers. • Tell time to the hour/half/quarter hour. • Draw a conclusion from bar graph, line graph, or pie chart.	• Complete a simple pattern. • Order three whole numbers (through 50). • Identify line graphs and a pie chart.	• Identify shapes. • Select the appropriate measuring tool. • Compare two whole numbers (10 or greater). • Identify simple valid arguments using *if…..then* statements. • Demonstrate number concepts using manipulatives, symbols, objects, or pictures. • Match numerals in contextual situations. • Identify/match whole numbers in contextual situations**.**
88 --- 73	72 --- 41	40 --- 21

These descriptors do not include all the skills and knowledge as contained in the Alternate Mathematics Standard.

Arizona Alternate Standard Performance Level Descriptors
Grade 4 Science

Exceeds the Standard – Students with significant cognitive disabilities who score in this level can typically function independently or with minimal cueing to demonstrate mastery of subject matter as reflected by the alternate science standard.

Meets the Standard – Students with significant cognitive disabilities who score in this level can typically function with moderate support through the use of visual representations, manipulatives, and objects to demonstrate a solid understanding of subject matter as reflected by the alternate science standard.

Approaches the Standard – Students with significant cognitive disabilities who score in this level can typically function with extensive support through the use of visual representations, manipulatives, and objects to demonstrate partial understanding of subject matter as reflected by the alternate science standard.
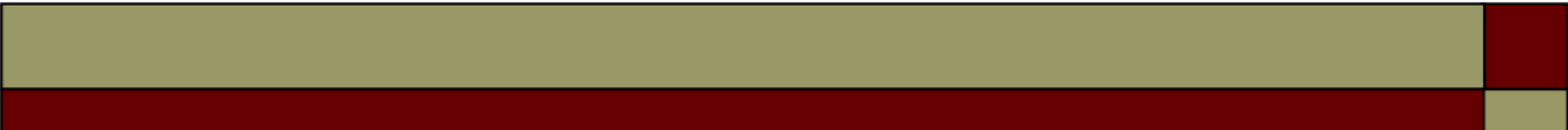
Falls Far Below the Standard – Students with significant cognitive disabilities who score in this level may have significant gaps and limited knowledge and skills that are necessary to satisfactorily meet the state's alternate science standard. Students will typically require a considerable amount of additional instruction and intervention in order to achieve a satisfactory level of understanding.

Students at the "**Exceeds the Standard**" level generally know the skills required at the "Meets" and "Approaches" levels and are able to:	Students at the "**Meets the Standard**" level generally know the skills required at the "Approaches" level and are able to:	Students at the "**Approaches the Standard**" level generally know and are able to:
• Identify seasons. • Use magnets with a variety of objects. • Identify a characteristic of an animal that helps it to survive.	• Select a resource that could be used in an investigation. • Communicate an observation. • Select technology that improves lives.	• Identify the sources of water. • Identify characteristic of an animal. • Identify science related career using pictures/manipulatives. • Demonstrate safe behavior when conducting an experiment. • Identify parts of a plant or animal. • Demonstrate how components of a system work.
80 --- 73	72-- 45	44--- 15

These descriptors do not include all the skills and knowledge as contained in the Alternate Science Standard.

Appendix D

Standard Setting Training Slides

2009 Standard Setting
for the
Arizona Alternate Assessment
(AIMS A)

Stephen N. Elliott, PhD
Vanderbilt University
Nashville, TN

Standard Setting Session Goals

1. Review all AIMS A items, current item difficulty (mean percent correct) data, and estimates of potential impact

2. Set Performance Level cut scores for the AIMS A using the Bookmark Procedure
   - Grades 3, 4, 5, 6, 7, 8, & 10 for Reading & Mathematics
   - Grades 4, 8, & 10 for Science

3. Provide feedback to standard setting panel on cut scores & refine AIMS A performance level descriptors.

4. Report to State Board of Education on May 18, 2009.

Key question to be answered: How much is enough?

Session Leader's Brief Bio

- PhD in Educational Psychology, Arizona State University (1980)

- Professor of Special Education and Dunn Family Chair of Educational & Psychological Assessment, Vanderbilt University

- Director, Learning Sciences Institute, Vanderbilt University

- Principal Investigator for 4 USDE projects concerning inclusive assessment design and practice; consultant on 4 other statewide projects (in AZ, ID, MS, SC) concerning the assessment of students with significant disabilities

- Author of 100+ articles and chapters on assessment of children with disabilities or at risk for educational difficulties.

- Led standard settings for Alternate Assessments in HI, ID, MS, WI, & AZ.

AZ Alternate Assessment & Data Management Leaders

ARIZONA DEPARTMENT OF EDUCATION

ADE Support Team

- Roberta Alley, Deputy Associate Superintendent
- Charles Bruen, Ed.D., Director of Data Analysis
- Danielle Gordon, Data Analysis and Technical Quality Coordinator
- Leila Williams, Ph.D. Alternate Assessment Coordinator
- Melanie Mosiman, Coordinator of AIMS EA
- Marilee Beach, Coordination of AIMS support materials
- Forster Okoli, Data Analyst

Standard Setting Session (3 day) Overview

- Introductions
- Workshop Goals & Roles of Participants
- Background of AIMS A Reading, Math, and Science
- Standard Setting Rationale & Bookmark Procedure
- Definitions of AIMS A Performance Levels
- Introduce the Major Steps in Bookmark Procedure
- Table Assignments & Decision Making Guidelines
- Review the AIMS A Items, Data & Scoring Criteria
- Review Standard Setting Procedures and Discuss Issues
- Recommend cut scores at each Grade for Reading, Math, & Science
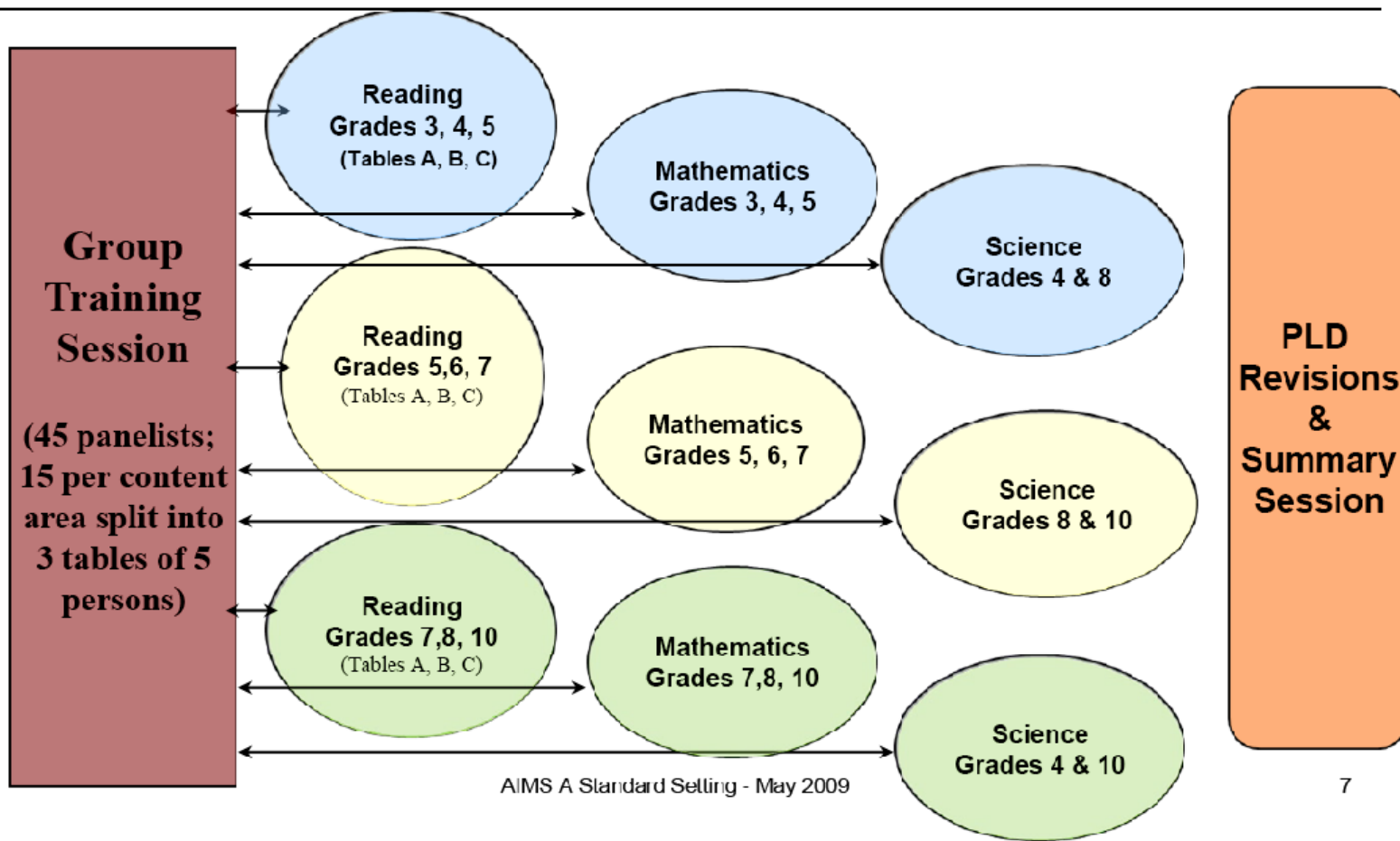- Review Results of Standard Setting for Each Content Area

Standard Setting Rationale:
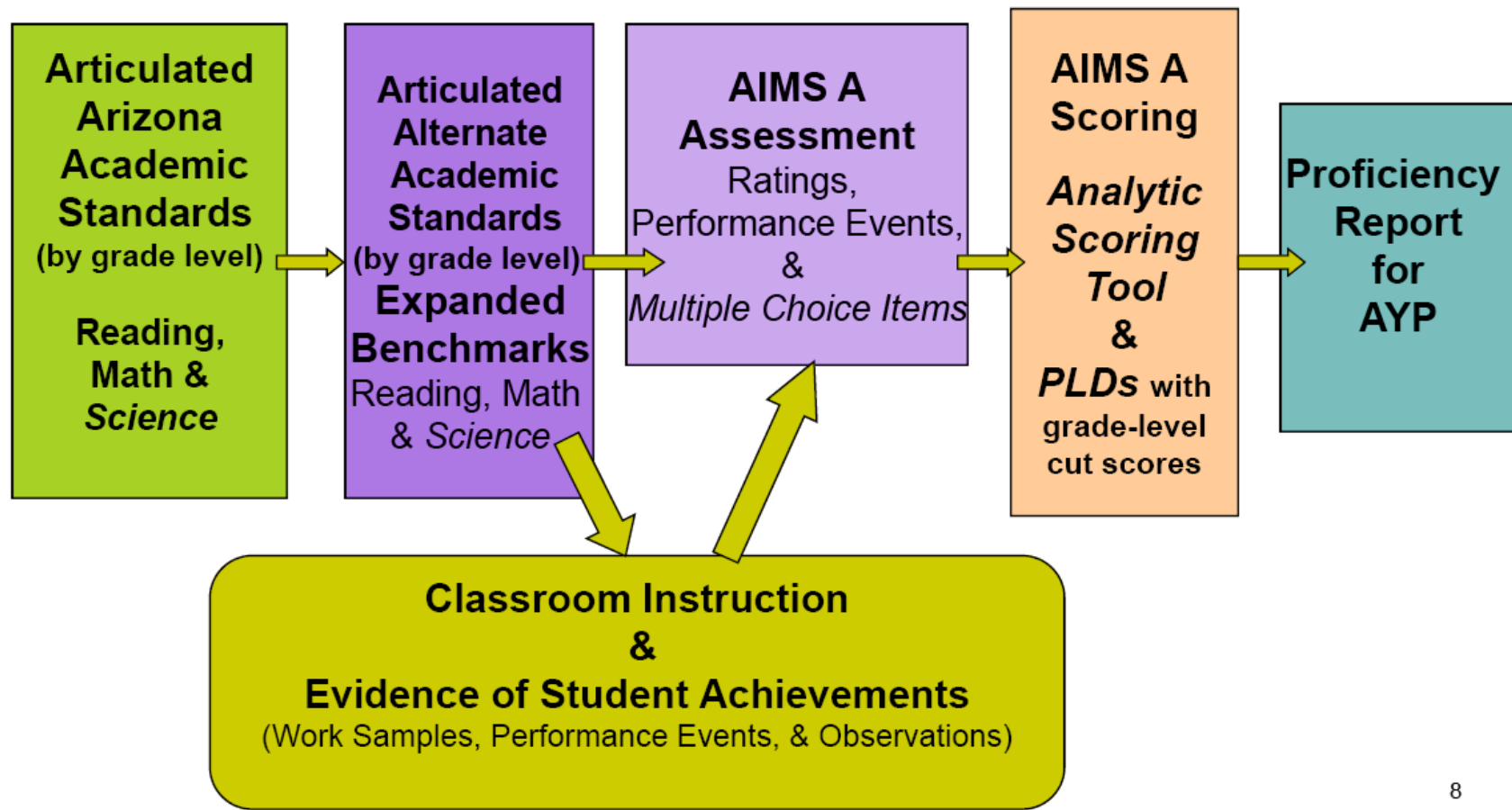# Establishing Alternate Achievement Standards

Judgment Based Approach

- ☐ Item Mapping Method (Bookmarking Procedure)

- ☐ A group of 45 stakeholders (teachers, administrators, content teachers, etc.) participate in a multi-day process that will result in <u>recommended</u> cut points on the AIMS A for Spring 2009

- ☐ Cut scores are based on what students in each performance level in each content area should know and be able to perform

2009 AIMS A Standard Setting Session:

Groups, Content, & Grades

Group Training Session

(45 panelists; 15 per content area split into 3 tables of 5 persons)

Reading Grades 3, 4, 5 (Tables A, B, C)

Mathematics Grades 3, 4, 5

Science Grades 4 & 8

Reading Grades 5, 6, 7 (Tables A, B, C)

Mathematics Grades 5, 6, 7

Science Grades 8 & 10

Reading Grades 7, 8, 10 (Tables A, B, C)

Mathematics Grades 7, 8, 10

Science Grades 4 & 10

PLD Revisions & Summary Session

AIMS A Standard Setting - May 2009

7

Review of the AIMS A Components

Articulated Arizona Academic Standards (by grade level) Reading, Math & *Science* → Articulated Alternate Academic Standards (by grade level) Expanded Benchmarks Reading, Math & *Science* → AIMS A Assessment Ratings, Performance Events, & *Multiple Choice Items* → AIMS A Scoring *Analytic Scoring Tool* & *PLDs* with grade-level cut scores → Proficiency Report for AYP

Classroom Instruction & Evidence of Student Achievements (Work Samples, Performance Events, & Observations)

8

Content Standards Assessed by AIMS A

- **Reading: 3 Strands (20 items at every grade level)**
  1. Reading Process
  2. Comprehending Literary Text
  3. Comprehending Informational Text

- **Mathematics: 5 Strands (22 items at every grade level)**
  1. Number Sense & Operations
  2. Data Analysis, Probability, and Discrete Math
  3. Patterns, Algebra, & Functions
  4. Measurement
  5. Structure & Logic

- **Science: 6 Strands (20 items at every grade level)**
  1. Inquiry Process
  2. History/Nature of Science
  3. Personal/Social Perspectives
  4. Life Science
  5. Physical Science
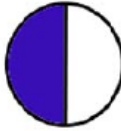 6. Earth/Space Science

Sample Multiple Choice:
6th Grade Reading

What is on September 22?

Sample Multiple Choice: 6th Grade Math

Which is the largest?

A: 1/4 B: 1/3 C: 1/2

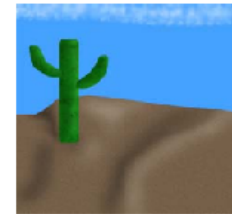# Sample Multiple Choice:
4th Grade Science
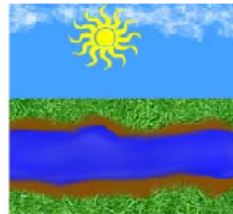
Sample Rater Item

RATER ITEMS
AIMS A
GRADE 5 Math

STUDENT NAME _____   DATE _____

TEACHER _____

Prompt Objective	Type of Assistance	Score
PRACTICE *"Pick 1."* Given 10 blocks, student picks 1.	PRACTICE	PRACTICE
1. *"What number is larger, 11 or 20?"* Student identifies the larger number, 11 or 20, using a number line.		
2. *"Pick the pie chart."* Student identifies the bar graph from a variety of graphic representations using pictures, symbols, text, manipulatives, or actions.		
3. *"What comes next in the pattern?"* Student adds to a pattern of 3 or more images/numbers using pictures, symbols, text, manipulatives, or actions.		
4. *"Record this data."* Student records given data for a probability activity.		
5. *"How many do you see?"* Student estimates a number of items presented using pictures, symbols, text, manipulatives, or actions.		

13

Rater Item Scoring Rubric

AIMS A			
RATER ITEM SCORING RUBRIC			
Level 4	**Level 3**	**Level 2**	**Level 1**
The student correctly performs the task without assistance or with a single repetition of instructions or refocusing through natural cues. Cues may include wait time or pointing.	The student correctly performs the task with general prompts and a single cue. Cues may include physical/verbal cues, auditory cues, objects, tactual cues, visual cues, or sign language.	The student correctly performs the task with specific prompts and up to 2 cues. Cues may include physical/verbal cues, auditory cues, objects, tactual cues, visual cues, or sign language.	The student does not perform the task at Level 2 or provides an incorrect response despite Level 2 support. Student requires extensive assistance and cannot perform the task without full adult support (hand over hand).
• The student responds or performs task correctly with no assistance. • If the student does not respond independently, responds incorrectly, or does not perform the requested task when given wait time, the teacher repeats the instructions and/or refocuses the student's attention.	• If the student responds incorrectly or does not perform the task at Level 4 when given wait time, the teacher provides general prompts and includes a single cue for the expected response from the student: 　o Elaborate or provide additional clarifying information on directions or expected response. 　o Demonstrate a similar response; "This is a picture of a dog. Show me the picture of a cat."	• If the student responds incorrectly or does not perform the task at Level 3 when given wait time, the teacher provides specific prompts and cues to direct the student's correct response: 　o Model exact response; "This is a picture of a dog. What is this?" (Show a picture/object representing a dog.) 　o Physically guide the student to the correct response.	
The student then responds correctly.	The student then responds correctly.	The student responds correctly *after* being given the correct answer.	The student does not respond or does not respond correctly. Teacher demonstrates response and moves on to the next prompt.
Record a score of 4	**Record a score of 3**	**Record a score of 2**	**Record a score of 1**
If the student still does not respond correctly– **move to Level 3 supports.**	If the student still does not respond correctly– **move to Level 2 supports.**	If the student still does not respond correctly– **move to Level 1 supports.**	

4 pts.　　　　　　　2 pts.　　　　　　　1 pt.　　　　　　　0 pts.

14

Performance Tasks Scoring

PERFORMANCE TASKS
AIMS A
GRADE 8 Reading

STUDENT NAME _____   DATE _____

TEACHER _____

Prompt Objective	0	1	2
PRACTICE – NO SCORE **Show story "Anna's First Day of School."** *"Point to the A in Anna."*	**PRACTICE – NO SCORE** The student is unable to perform the task.	**PRACTICE – NO SCORE** The student is able to point to A after the teacher models the correct response.	**PRACTICE – NO SCORE** The student points to A without assistance or with a single repetition of instruction or redirection.
1.1 Read story "Anna's First Day of School." *"Who is this story about?"*	The student is unable to perform the task.	The student is able to indicate Anna after the teacher models the correct response.	The student indicates Anna *without* assistance or with a single repetition of instruction or redirection.
1.2 Reread story "Anna's First Day of School" if necessary and show the picture cards of Anna at home and Anna in school. *"Where is Anna going?"*	The student is unable to perform the task.	The student is able to indicate school after the teacher models the correct response.	The student indicates school without assistance or with a single repetition of instruction or redirection.
1.3 Reread story "Anna's First Day of School" if necessary and show the map and the dictionary. *"What does Anna use to find English class?"*	The student is unable to perform the task.	The student is able to indicate a map after the teacher models the correct response.	The student indicates a map without assistance or with a single repetition of instruction or redirection.
1.4 Reread story "Anna's First Day of School" if necessary and show the map. *"What class does Anna have after English?"*	The student is unable to perform the task.	The student is able to indicate on the schedule after the teacher models the correct response.	The student indicates on the schedule without assistance or with a single repetition of instruction or redirection.
1.5 Reread story "Anna's First Day of School" if necessary and show the word cards big and small. *"How does Anna feel about school?"*	The student is unable to perform the task.	The student is able to indicate emotion after the teacher models the correct response.	The student indicates emotion without assistance or with a single repetition of instruction or redirection.

0 pts.   2 pts.   4 pts.

15

Item Scoring Summary

- Each multiple-choice item is scored 0 or 4
- Each performance item is scored 0, 2, or 4
- Each rating item is scored 0, 1, 2, or 4

Thus, regardless of the type of item or content area, a score of 0 mean "cannot do" and a score of 4 "can do without any assistance." The result is…

- Reading total scores ranging from 0 to 80
- Math total scores ranging from 0 to 88
- Science total scores ranging from 0 to 80

16

Mean Score Data Across Grades

AIMS A Total Mean Scores

	Grade 3 Mean (SD) # Items Mean % Total Score	Grade 4 Mean (SD) # Items Mean % Total Score	Grade 5 Mean (SD) # Items Mean % Total Score	Grade 6 Mean (SD) # Items Mean % Total Score	Grade 7 Mean (SD) # Items Mean % Total Score	Grade 8 Mean (SD) # Items Mean % Total Score	Grade 10 Mean (SD) # Items Mean % Total Score
Reading	.6166 (.13) 20 61.66% 49.33	.6604 (.11) 20 66.04% 52.83	.6350 (.10) 20 63.50% 50.80	.6194 (.10) 20 61.94% 49.55	.6461 (.10) 20 64.61% 51.69	.6376 (.11) 20 63.76% 51.01	.6901 (.11) 20 69.01% 55.21
Math	.6274 (.15) 22 62.74% 55.21	.6211 (.14) 22 62.11% 54.66	.5623 (.16) 22 56.23% 49.48	.5822 (.12) 22 58.22% 51.23	.5989 (.11) 22 59.89% 52.70	.5871 (.12) 22 58.71% 51.66	.5735 (.10) 22 57.35% 50.47
Science		.6816 (.10) 20 68.16% 59.98				.7386 (.08) 20 73.86% 65.00	.6875 (.13) 20 68.75% 60.5

17

Transforming AIMS A Scores

- To facilitate comparisons of total scores on AIMS A where different tests or subscales that have different numbers of items (e.g., 20 Reading items, 22 Math items), we use percentage correct scores. These scores are then transformed mathematically to an individual Reading, Math, or Science total score based on the total possible number of points earned. The final transformation of scores to a performance level for AYP reporting is done by a standard setting panel and is based on their consensus professional judgment.

- The table below provides examples of AIMS A Reading and Math score transformations. Given the Science test has 22 items, the transformations for it are the same as the Math Test.

Reading % Correct Score	Reading Total Score	Reading Performance Level		Math % Correct Score	Math Total Score	Math Performance Level
0	0	?		0	0	?
.10	8	?		.10	8.8	?
.20	16	?		.20	17.6	?
.30	24	?		.30	26.4	?
.40	32	?		.40	35.2	?
.50	40	?		.50	44	?
.60	48	?		.60	52.8	?
.70	56	?		.70	61.6	?
.80	64	?		.80	70.4	?
.90	72	?		.90	79.2	?
1.0	80	?		1.0	88	?

18

Score Variability & Confidence Bands

☐ The **mean score** is the most representative score for a group, however, when scores **vary considerably** one must be cautious about using the mean to make important decisions.

☐ A **confidence band** is used in statistical analysis to represent the uncertainty in an estimate of a curve or function based on limited or noisy data. Confidence bands are often used as part of the graphical presentation of results in a statistical analysis. <u>Confidence bands represent the uncertainty in an estimate of a single numerical value.</u>

19

Item Score Distribution & Confidence Bands

Item #8 has a mean of approx. 3.25 with a standard deviation of .10. This results in a CB of approx. .20. As you can see, this item is the easiest reading items.

The hardest item is #22. It has a mean of approx. 1.5 and less variability than all other items.

The data in this graphic is based on 826 4th grade students.

Grade 4 Reading

Item Number

8 19 12 1 10 11 6 7 9 20 18 3 23 17 13 16 21 25 24 22

Mean Score

Legend

■ RI
▲ PT
◆ MC

N=826

20

Interpreting Scores:
4 Level Performance Descriptors

Students earn a Total Score for each content area. The total scores are used to guide the determination of which of the four Performance Levels best describe the students' achievement.

Falls Far Below → **Approaches** → **Meets** → **Exceeds**
the Standard   the Standard   the Standard   the Standard

The translation of a Total Score to a Performance Level is a **professional judgment!**

Excellent judgments are based on a clear understanding of what is expected of the learner, what the assessment measures, and how the group actually performed on the assessment.

Example AIMS A PLD: Grade 4 Reading

Exceeds the Standard – Students with significant cognitive disabilities who score in this level can typically function independently or with minimal cueing to demonstrate mastery of subject matter as reflected by the alternate reading standard.

Meets the Standard – Students with significant cognitive disabilities who score in this level can typically function with moderate support through the use of visual representations, manipulatives, and objects to demonstrate a solid understanding of subject matter as reflected by the alternate reading standard.

Approaches the Standard – Students with significant cognitive disabilities who score in this level can typically function with extensive support through the use of visual representations, manipulatives, and objects to demonstrate partial understanding of subject matter as reflected by the alternate reading standard.

Falls Far Below the Standard – Students with significant cognitive disabilities who score in this level may have significant gaps and limited knowledge and skills that are necessary to satisfactorily meet the state's alternate reading standard. Students will typically require a considerable amount of additional instruction and intervention in order to achieve a satisfactory level of understanding.

Reading Gr 4 PLD with Specific Skills for Exceeds, Meets, & Approaches the Standard

Students at the "Exceeds the Standard" level generally know the skills required at the "Meets" and "Approaches" levels and are able to:	Students at the "Meets the Standard" level generally know the skills required at the "Approaches" level and are able to:	Students at the "Approaches the Standard" level generally know and are able to:	Students at the "Fall Far Below the Standard" level generally know and are able to:
• Follow a set of multi-step directions in order. • Identify specific facts in text. • Select a synonym, antonym and homonym. • Make a prediction.	• Determine meaning of a simple or environmental word. • Identify the conflict.	• Find a solution to a problem. • Identify one aspect of the setting.	• Pick one trait of a character.

These descriptors do not include all the skills and knowledge as contained in the Reading Standard.

Focus on Meets the Standard (Proficient)
4th grade

Students at the "Meets the Standard" level generally know the skills required at the "Approaches" level and are able to:

* Determine meaning of a simple or environmental word.
* Identify the context.

☐ **Meets the Standard** – Students with significant cognitive disabilities who score in this level can typically function with moderate support through the use of visual representations, manipulatives, and objects to demonstrate a solid understanding of subject matter as reflected by the alternate reading standard.

Four Performance Levels: Three Cuts Determining Performance Standards

Possible Raw Score Ranges

Reading 0---80
Math   0---88
Science 0---80

Low   High

Far Below Standard | Approaches Standard | Meets Standard | Exceeds Standard

Approaches Cut Score   **Meets Cut Score**   Exceeds Cut Score
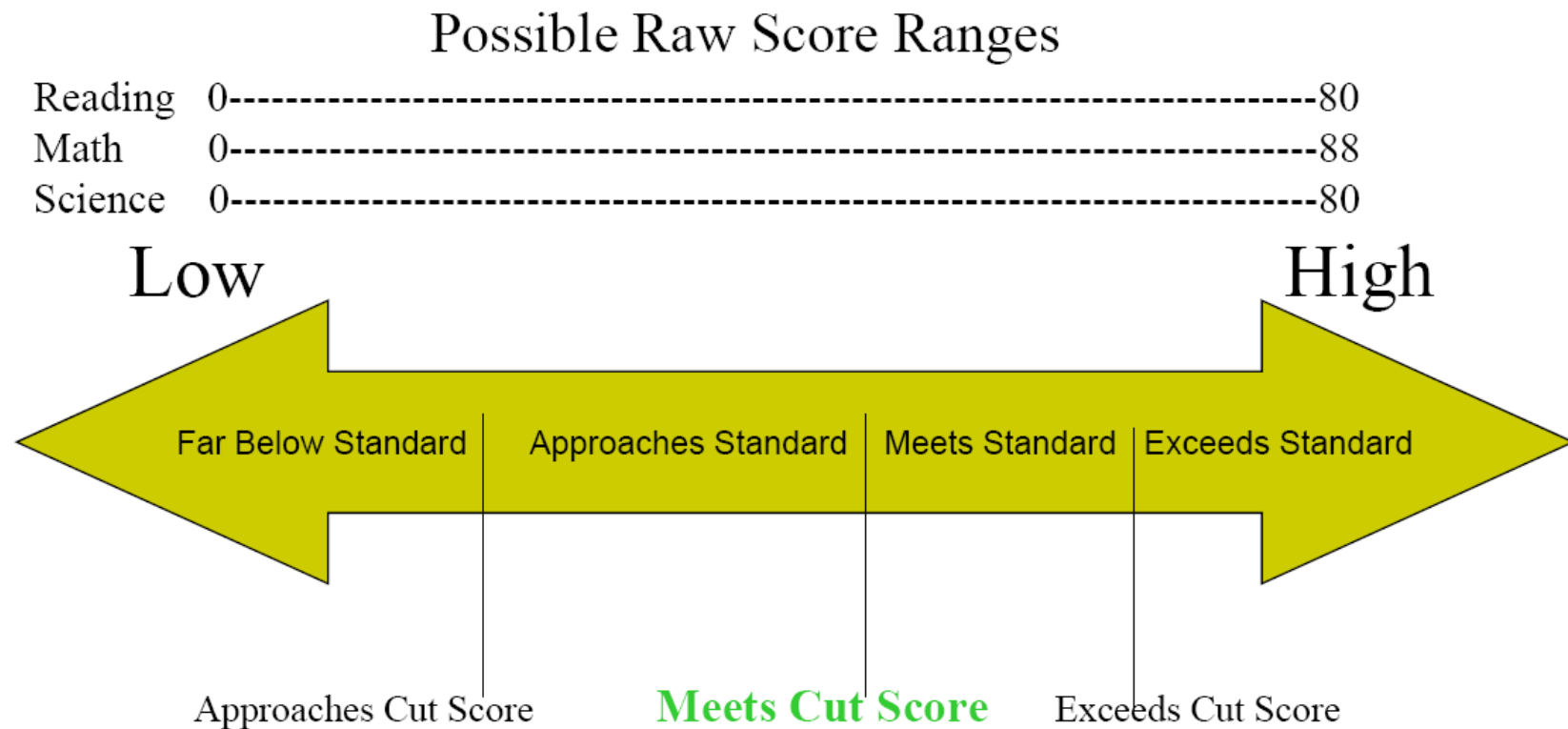
The Marginally Proficient Student (At the Threshold of Meets Standard)

□ Our task is to describe, in as much detail possible, how the <u>marginally proficient</u> student taking AIMS A would perform on each test item.

□ Discuss with your group what "Marginally Proficient" means in each content area. Remember to use the PLDs to help you refine a definition.

Bookmarking Procedure

☐ Participants receive a Booklet (Item Map) with a set of test items ordered from easiest to most difficult based on item statistics (mean % correct; the higher the percent correct, the easier the item).

☐ Participants study the items and determine the cut score by placing a bookmark (physical sheet or mark) at the location in the booklet where they think a student who is functioning at the Meets Standard level should likely perform.

☐ Items preceding the bookmark represent items that "proficient" students should likely perform.

The Marginally Proficient Student (At the Threshold of Meets Standard)

□ Panelists' task is to describe, in as much detail possible, how the marginally proficient (Meets Standard) student taking AIMS A would perform on each test item.

□ Think of Marginally Proficient as a student receiving special education services who is **<u>just</u> <u>demonstrating</u>** the knowledge and skills that s/he would be expected, based on the definition of Meets Standard, to show for each **grade**.

Sample OIB MAP for Grade 4 Reading

Grade 4 Reading Item Map

Item #8 is a multiple-choice item and is the easiest one on the gr. 4 reading assessment with a mean score of 3.27 and a p value of .8181.
Examine this item. What does it measure?

Which item is the 4th easiest item on the gr. 4 Reading Test? Examine this item. What makes it a little harder than #8?

OIB Page Number	AZID Number	Test Item Number	Mean Score	P-Value	Item Type	Score Key	Strand/ Concept/ PO	Why is this item more difficult that the last item(s)?
1	62094010	8	3.27	0.8181	MC	B	S2C1PO2	
2	62094104	19	3.22	0.8060	PT		S3C2PO1	
3	62094020	12	3.21	0.8036	MC	B	S2C1PO7	
4	62094030	1	3.19	0.7975	MC	C	S2C1PO5	

AIMS A Standard Setting - May 2009

29

Continuation of Sample Item Map

OIB Page Number	AZID Number	Test Item Number	Mean Score	P-Value	Item Type	Score Key	Strand/ Concept/ PO	Why is this item more difficult that the last item(s)?
15	62094002	13	2.38	0.5951	MC	C	S2C1PO5	
16	62094101	16	2.29	0.5733	PT		S3C2PO1	
17	62094201	21	2.24	0.5621	RI		S1C6PO1	
18	62094205	25	2.16	0.5418	RI		S3C2PO2	
19	62094204	24	2.12	0.5312	RI		S3C2PO2	
20	62094202	22	1.48	0.3709	RI		S1C4PO6	

The hardest grade 4 Reading Item, #22, is a Rating item and has a mean score of 1.48 and a p value of .3709.
Examine this item and discuss why you think it is the most difficult.

30

Cumulative Score Distributions: Impact Data

☐ Before finalizing cut scores, panelist are encouraged to consider the likely effect or impact of them on students.

☐ By looking at the cumulative distribution of total scores – from 0 to 80 – one can determine the percentage of students who would likely be above and below each cut point.

Grade 4 Reading

Raw Score	Frequency	Percent	Cumulative Percent
0	80	8.91%	8.91%
1	3	0.33%	9.24%
2	2	0.22%	9.47%
3	1	0.11%	9.58%
4	6	0.67%	10.24%
5	2	0.22%	10.47%
6	2	0.22%	10.69%
7	2	0.22%	10.91%
8	3	0.33%	11.25%
9	0	0.00%	11.25%
10	2	0.22%	11.47%
72	35	3.90%	85.52%
73	14	1.56%	87.08%
74	25	2.78%	89.87%
75	14	1.56%	91.43%
76	37	4.12%	95.55%
77	5	0.56%	96.10%
78	17	1.89%	98.00%
79	0	0.00%	98.00%
80	18	2.00%	100.00%

Additional Descriptive Statistics

Grade 4 Reading

□ Along with the
cumulative frequency
distributions and
percentage of students
with each score, you
also have common
descriptive statistics for
each grade level test.

Statistics	
N	898
Mean	48.53
Median	55
Mode	0
Std. Deviation	23.94
Percentile 25	33
50	55
75	68

32

Activity: Connect "Meets the Standard" PLD for Reading to the Item Data

- Step 1. Re-read the definition of Meets the Standard for Reading at one of your grade level. Note the defining knowledge & skills listed.

- Step 2. Examine the Reading items at one of your grade levels. Try to find one or more items that represent the defining knowledge & skills for Meets the Standard.

- Step 3. What are the Mean Scores for the items you located? What makes these items more difficult than others located above it in the Item Map?

- Step 4. Should students who Meet the Standard be expected to do well on these items? What percent of the students Meeting the Standard would you find acceptable?

Major Steps in Bookmarking Procedure
For Grade Performance Level Cut Scores

- **Round 1**: <u>Individual</u> & Performance cut score

- Post-Round #1 Discussion

- **Round 2:** <u>Team Consensus</u> for Performance cut score

- Post-Round #2 Discussion with feedback on impact

- **Round 3:** <u>Teams Final Decisions</u>

- Post Round #3: Feedback on Median cut score & likely impact on student distributions

Informed Judgments: Key Steps & Resources

Standard setting is predicated on informed
judgments by knowledgeable panelists.

| Read PLDs | → | Review Item Map & Items | → | Connect Defining Items & Data | → | Discuss & Decide Easiest Item for Level |

Decision Making Guidelines

- ☐ Professional Judgments
- ☐ Tolerance for Different Judgments
- ☐ Consensus Building Process
- ☐ Decision-Making Teams or Tables should be Representative
- ☐ Decision-Making Teams need a Leader
- ☐ No Right or Wrong Answers
- ☐ The Resulting Performance Standards are <u>Advisory</u>

Round 1 Form for Meets the Standard Decision

Group # ____ Grade: ____ Content Area: _____ Date: _____

Meets Member	Round 1: Individual Recommended Cut Point For Meets	Round 2: Consensus Recommended Cut Point For Meets	Round 3: Final Consensus Recommended Cut Point For Meets
A			
B			
C			
D			
E			
F			
Group	Median		

Each Member of the Group Decides on a Cut Score Independently and the Group Leader Lists it down

Group Leader Finds the Median by Finding the Middle value or Averaging the Two Middle Numbers in the Group

Calculating the Median Score

☐ The **median** is described as the number separating the higher half of a sample or a population from the lower half.

☐ The **median** of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of observations, the median is not unique, so one often takes the <u>mean</u> of the two middle values. At most half the population have values less than the **median** and at most half have values greater than the median. If both groups contain less than half the population, then some of the population is exactly equal to the median. For example, if $a < b < c$, then the median of the list $\{a, b, c\}$ is b, and if $a < b < c < d$, then the median of the list $\{a, b, c, d\}$ is the mean of b and c, i.e. it is $(b + c)/2$.

☐ The median can be used when a distribution is skewed, when end values are not known, or when outliers likely represent measurement errors.

Round 2 Form for Meets the Standard Decision

Group # ____ Grade: ____ Content Area: _____ Date: _____

Meets Member	Round 1: Individual Recommended Cut Point For Meets	Round 2: Consensus Recommended Cut Point For Meets	Round 3: Final Consensus Recommended Cut Point For Meets
A			
B			
C			
D			
E			
F			
Group	Median	One Score	One Score

Group Consensus After Discussion and Sharing of Individual Scores and Median

39

Round 3 Form for "Meets the Standard" Decision

Group # _____ Grade: _____ Content Area: _____ Date: _____

Meets Member	Round 1: Individual Recommended Cut Score For Meets Standard	Round 2: Consensus Recommended Cut Score For Meets Standard	Round 3: Final Consensus Recommended Cut Score For Meets Standard
A			
B			
C			
D			
E			
Group	Median	One Score	One Score

One Final Group Consensus Score After Looking At Impact Data & Discussion

Group Leader: _____
Signature

Procedure for Approaches & Exceeds Standards Cut Score Decisions

☐ Only Round 2 with Impact data for these levels.

☐ We will find the median of the scores from all groups to get the Approaches & Exceeds Cut score.

Approaches Member	Round 2: Consensus Recommended Cut Point For Approaches
A	
B	
C	
D	
E	
Group	Median

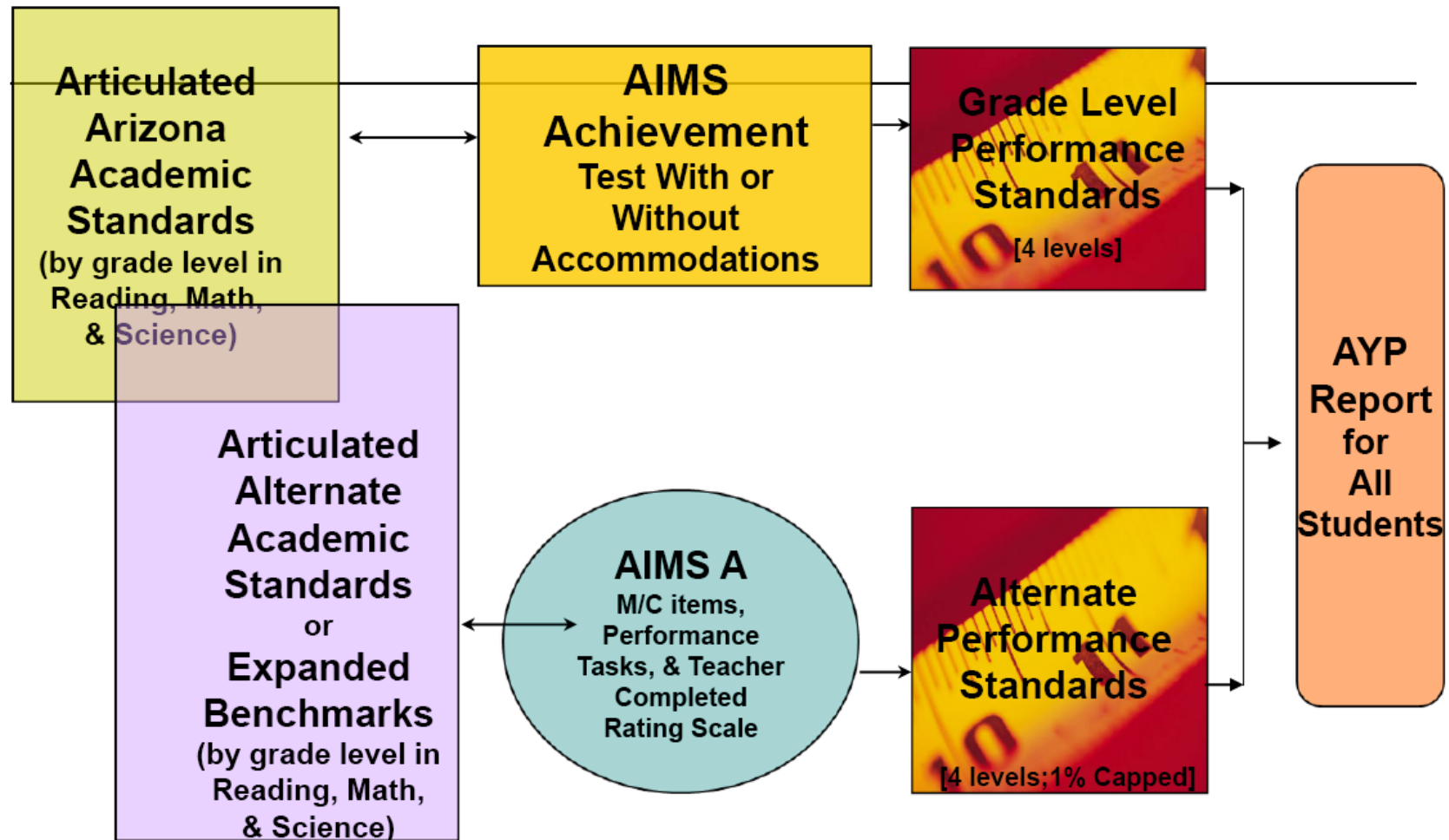| Exceeds Member	Round 2: Consensus Recommended Cut Point For Exceeds
A	
B	
C	
D	
E	
Group	Median

AIMS A Standard Setting - May 2009

41

Review Standard Setting Procedures, Discuss Any Concerns, & Refine PLDs

☐ What student "Should" know versus what they "Do" know

☐ What knowledge, skills and abilities separate:

  - Falls Far Below Standards from Approaches Standards

  - **Approaches Standards from Meets Standards**

  - Meets Standards from Exceeds Standards

☐ Think about students at the threshold of each level

☐ All AIMS A Students – **not just <u>your</u> students**

☐ Refine/update PLDs to include specific examples of skills; keep notes on issues or concerns to facilitate revision work.

Outcome: An Integrated Arizona Assessment System

Articulated Arizona Academic Standards
(by grade level in Reading, Math, & Science)

AIMS Achievement Test With or Without Accommodations

Grade Level Performance Standards
[4 levels]

Articulated Alternate Academic Standards
or
Expanded Benchmarks
(by grade level in Reading, Math, & Science)

AIMS A
M/C items, Performance Tasks, & Teacher Completed Rating Scale

Alternate Performance Standards
[4 levels; 1% Capped]

AYP Report for All Students

AIMS A Standard Setting - May 2009

43

Thank you for the opportunity to work with you to determine AIMS A Performance Standards!

Contact Information:

Stephen N. Elliott, PhD

steve.elliott@vanderbilt.edu

615-322-2538

Appendix E

Sample Item Map

AIMS A Bookmark Standard Setting May 2009

Grade 4 Reading Item Map

OIB Page Number	AZID Number	Test Item Number	Mean Score	P-Value	Item Type	Score Key	Strand/ Concept/ PO	Why is this item more difficult that the last item(s)?
1	62094010	8	3.27	0.8181	MC	B	S2C1PO2	
2	62094104	19	3.22	0.8060	PT		S3C2PO1	
3	62094020	12	3.21	0.8036	MC	B	S2C1PO7	
4	62094030	1	3.19	0.7975	MC	C	S2C1PO5	
5	62094004	10	3.15	0.7878	MC	A	S2C1PO7	
6	62094009	11	2.86	0.7163	MC	B	S2C1PO2	
7	62094023	6	2.85	0.7127	MC	C	S2C1PO7	
8	62094011	7	2.79	0.6993	MC	A	S2C1PO2	
9	62094003	9	2.75	0.6896	MC	C	S2C1PO2	
10	62094105	20	2.71	0.6781	PT		S3C1PO7	
11	62094103	18	2.69	0.6733	PT		S3C2PO1	
12	62094032	3	2.53	0.6339	MC	A	S1C4PO5	
13	62094203	23	2.47	0.6184	RI		S3C2PO1	
14	62094102	17	2.39	0.5981	PT		S3C2PO1	
15	62094002	13	2.38	0.5951	MC	C	S2C1PO5	
16	62094101	16	2.29	0.5733	PT		S3C2PO1	
17	62094201	21	2.24	0.5621	RI		S1C6PO1	
18	62094205	25	2.16	0.5418	RI		S3C2PO2	
19	62094204	24	2.12	0.5312	RI		S3C2PO2	
20	62094202	22	1.48	0.3709	RI		S1C4PO6	

Appendix F

Sample Item Distribution Graph

AIMS A Standard Setting / Final Summary Report / May 2009
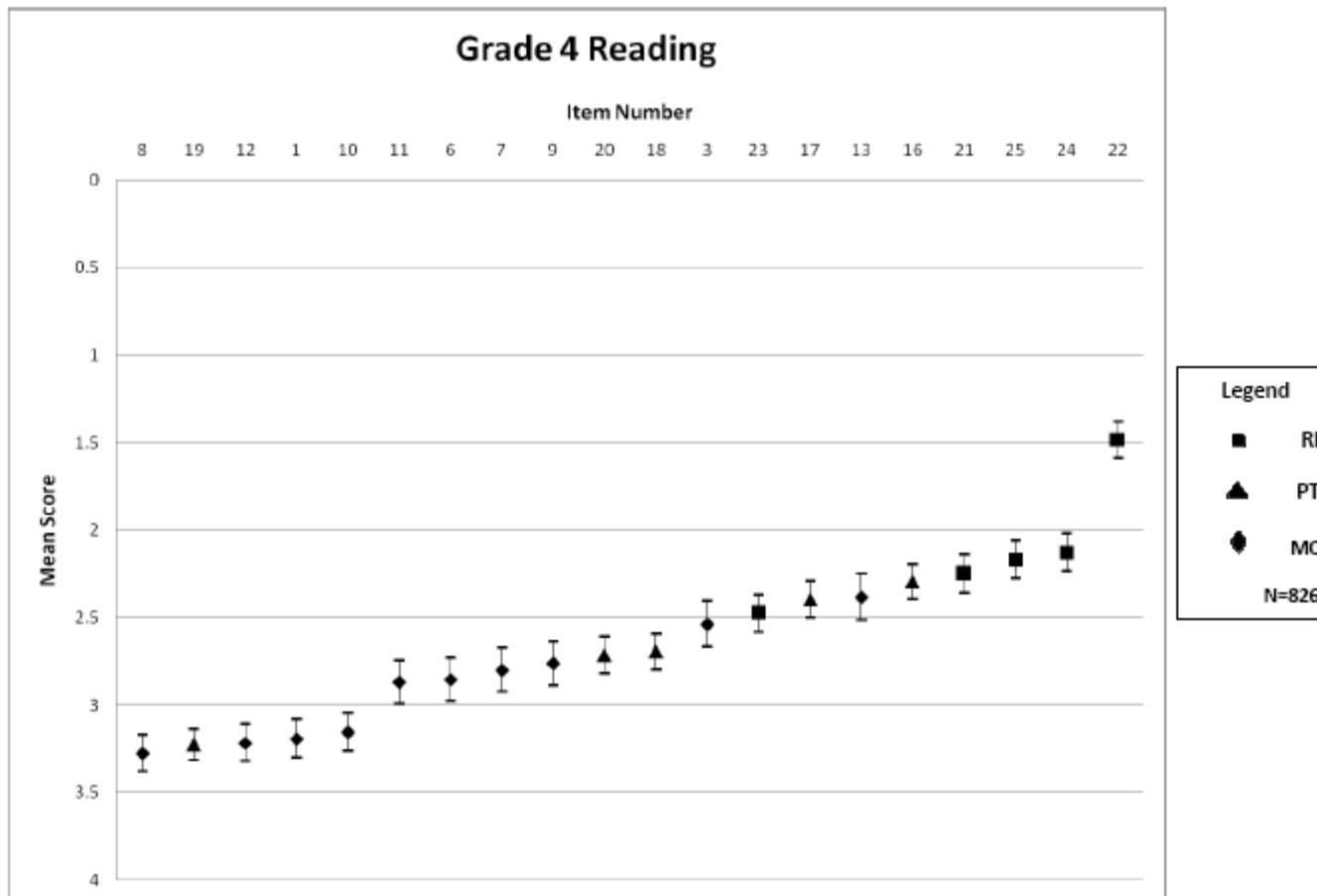
Appendix G

Sample Cumulative Score Distribution for Impact Analysis

Grade 4 Reading

Statistics	
N	898
Mean	48.53
Median	55
Mode	0
Std. Deviation	23.94
Percentile 25	33
50	55
75	68

Raw Score	Frequency	Percent	Cumulative Percent
0	80	9%	9%
1	3	0%	9%
2	2	0%	10%
3	1	0%	10%
4	6	1%	10%
5	2	0%	11%
6	2	0%	11%
7	2	0%	11%
8	3	0%	11%
9	0	0%	11%
10	2	0%	12%
11	4	0%	12%
12	3	0%	12%
13	7	1%	13%
14	2	0%	13%
15	2	0%	14%
16	4	0%	14%
17	2	0%	14%
18	5	1%	15%
19	4	0%	15%
20	8	1%	16%
21	3	0%	16%
22	0	0%	16%
23	5	1%	17%
24	7	1%	18%
25	9	1%	19%
26	6	1%	19%
27	6	1%	20%
28	7	1%	21%
29	5	1%	21%
30	9	1%	22%
31	7	1%	23%
32	14	2%	25%
33	10	1%	26%
34	8	1%	27%
35	8	1%	28%
36	12	1%	29%
37	10	1%	30%
38	11	1%	31%

39	6	1%	32%
40	6	1%	33%
41	3	0%	33%
42	8	1%	34%
43	9	1%	35%
44	12	1%	36%
45	5	1%	37%
46	14	2%	38%
47	12	1%	40%
48	18	2%	42%
49	9	1%	43%
50	14	2%	44%
51	13	1%	46%
52	11	1%	47%
53	16	2%	49%
54	10	1%	50%
55	10	1%	51%
56	12	1%	52%
57	17	2%	54%
58	6	1%	55%
59	13	1%	56%
60	24	3%	59%
61	16	2%	61%
62	16	2%	63%
63	15	2%	64%
64	30	3%	68%
65	13	1%	69%
66	19	2%	71%
67	23	3%	74%
68	26	3%	77%
69	17	2%	78%
70	15	2%	80%
71	17	2%	82%
72	35	4%	86%
73	14	2%	88%
74	25	3%	90%
75	14	2%	92%
76	37	4%	96%
77	5	1%	97%
78	17	2%	98%
79	0	0%	98%
80	18	2%	100%