2010 NeSA Reading Standard Setting Technical Report

June 28-30, 2010

Prepared by
Data Recognition Corporation

DATA RECOGNITION

DRC

CORPORATION

TABLE OF CONTENTS

Section 1: Executive Summary **1**

Section 2: Introduction **3**

2.1 Background 3

2.2 Purpose and Objectives of NeSA and Standard Setting Event 3

2.3 Bookmark Standard Setting Method 4

2.4 Contrasting Groups Standard Setting Method 5

2.5 Meeting with a Committee of Stakeholders 5

Section 3: Preparation for Standard Setting **6**

3.1 Bookmark Panelist Recruitment 6

3.2 Roles and Responsibilities 7

3.3 Materials Preparation 8

3.4 Ordered Item Booklet Item Placements 8

3.5 Ordered Item Booklet Preparation 9

Section 4: Standard Setting Procedures **10**

4.1 Contrasting Group Procedure 10

4.2 Modified Bookmark Procedure 13

4.3 Vertical Articulation Across Grades 14

4.4 Merging Bookmark and Contrasting Groups 15

Section 5: Results **16**

5.1 Contrasting Groups Analyses 16

5.2 Bookmark Analyses 17

5.3 Recommendation and Approval of State Board of Education 18

5.4 Panelists' Survey Evaluation Results 19

References **20**

Appendices:

A. NeSA-R Performance Level Descriptors 21

B. Meeting Agenda 42

C. PowerPoint: Setting Academic Proficiency Standards 45

D. Impacts by Round 55

E. Item Separation Maps 56

F. Contrasting Groups Summaries 60
G. Contrasting Groups Analyses 70
H. Cut Scores and Impacts by Method 77
I. Panelist Evaluation Form 78
J. Bookmark Panelist Evaluation Summary 80
K. Cut Scores and Standard Errors of Measurement by Round……..… 81

1. Executive Summary

Establishing the academic performance levels for the NeSA-R involved a series of four events. A meeting including Nebraska State Board of Education (SBE) members and other stakeholders was held February 25, 2010 to familiarize them with the process and obtain their feedback to ensure the most effective and valid outcome possible. A contrasting groups survey of reading teachers and specialists was conducted in spring 2010, before the first operational assessment, to determine the overall proficiency level of Nebraska students, independent of a particular assessment. A formal Bookmark standard setting meeting was held after operational data were available, which was deemed the *method of record* for a recommendation to the SBE. Finally, the SBE met in early July to review the findings and to formally establish the performance levels. This report documents the Bookmark and Contrasting Groups events.

The Bookmarking event to set academic performance level cut scores for grades 3 through 8 and 11 in reading for the Nebraska Student Assessment (NeSA-R) was held on June 28-30, 2010 in Lincoln, Nebraska. The purpose of the meeting was to recommend cut scores that will be used to place students into three performance levels: *Below the Standards, Meets the Standards, Exceeds the Standards*. The final decision on cut scores was made by the State Board of Education July 7-8, 2010. The performance levels will be utilized by local, state, and federal accountability programs. The Meets the Standards and Exceeds the Standards levels are used for the *No Child Left Behind (NCLB)* Adequate Yearly Progress (AYP) proficiency goal, which requires annual progress in the percents of students falling into the Meets the Standards category or above.

One hundred and one educational stakeholders from Nebraska participated in the meetings. Committee members were selected to represent grades 3 through 8, high school, and higher education. The standard setting method known as the *Bookmark* procedure (Lewis, Mitzel, & Green, 1996) was employed. This approach was augmented by a *Contrasting Groups* survey of Nebraska teachers conducted shortly before the spring operational NeSA-R administration.

Bookmark is an *item-based* method that asks panelists to determine which items can be successfully answered (67% likelihood) by students at the performance level boundaries. Contrasting Groups is a *student-based* method that asks teachers to place students that they know into one of the three performance levels without considering the assessment per se. The success of either approach requires a shared understanding of what skills and knowledge are required at each level. This shared understanding is expressed in *Performance Level Descriptors* (PLD's).

The item-based Bookmark method is, perhaps, the most philosophically consistent method to use with criterion-reference, standards-based[1] assessments like the NeSA and was designated the method of record. In the course of the Bookmark process, panelists were shown results of the Contrasting Groups survey, impact data (percent of spring operational students in each performance level), and relevant

[1] It is somewhat unfortunate the term *standard* is used in two different senses in this area. *Content standards* are written descriptions of the goals and expectations for learning and instruction at each grade level. *Performance standards*, which are the focus of this report, define the levels of achievement necessary for each performance level. In some contexts, the term *performance standard* is interchangeable with *cut score*.

results from NAEP (National Assessment of Educational Progress) and the ACT college entrance exam. The State Board of Education (SBE) reviewed the results from both the Bookmark and Contrasting Groups studies. DRC presented another option of a simple, unweighted averaging of the logit cut points from the two studies. The average was computed in the logit metric and translated into percent of students in category. The percent in categories was not the statistic of focus; these were calculated after the logit cuts were determined.

Two notable adjustments were made to the option to arrive at the final cut scores:

1) Grade 8 was adjusted in "Exceeds the Standard" from 27.4 percent to 22.2 percent to more closely match the other grades, and,
2) All grades except grade 7 were adjusted to allow more Below the Standards students in the category and correspondingly fewer students in the Meets the Standards category.

Board-Approved Cut Scores

The final SBE approved cut scores and the percent of spring 2010 students expected to be in each performance level are shown in Table 1.1.1. Psychometrically, cut scores are defined in a logit metric, which are transformed percent correct scores. Logits are preferable to percent correct because they are not tied to a specific test form and thus will not change from year to year. This ensures a consistent definition of the performance levels even if different test forms vary somewhat in difficulty.

For reporting purposes, logits are converted into the Scale Score metric, which is mathematically equivalent but more user-friendly. The SBE determined that the Meets the Standards level will begin at a Scale Score of **85** for all grades, and the Exceeds the Standards will begin at **135**. These values will be used for all grades and will not change from year to year.

After items have been chosen for a form, the logit cut scores can be used to determine the raw-score cut points specific to that form.

Table 1.1.1 includes the logit cut scores, the 2010 Raw Score ranges for each performance level, the Scale Score, and the percent of spring 2010 students falling into each level. The logit and Scale Score values will not change in the future, but the raw score ranges may shift slightly to reflect any variation in item and form difficulty. The percent of students in each level is also expected to change to reflect improvement in student proficiency.

Table 1.1.1 State Board of Education Approved Standard Setting Results

	Logit Cut points		2010 Raw Score Ranges by Performance Level			Scale Score Ranges by Performance Level			2010 Percent in Each Performance Level		
Grade	Bel/Mt	Mt/Ex	Below	Meets	Exceeds	Below	Meets	Exceeds	Below	Meets	Exceeds
3	-0.5168	1.2340	0 to 29	30 to 40	41 to 45	1 to 84	85-134	135 to 200	32.5	47.4	20.1
4	-0.5117	0.8591	0 to 29	30 to 39	40 to 45	1 to 84	85-134	135 to 200	30.5	48.1	21.4
5	-0.4122	0.8560	0 to 31	32 to 41	42 to 48	1 to 84	85-134	135 to 200	32.6	48.2	19.2
6	-0.4331	0.8924	0 to 32	33 to 42	43 to 48	1 to 84	85-134	135 to 200	31.8	48.6	19.6
7	-0.5104	0.7855	0 to 29	30 to 40	41 to 48	1 to 84	85-134	135 to 200	31.0	48.0	21.0
8	-0.4812	0.8712	0 to 32	32 to 42	43 to 50	1 to 84	85-134	135 to 200	29.6	48.1	22.3
11	-0.4103	0.8508	0 to 31	32 to 42	43 to 50	1 to 84	85-134	135 to 200	31.5	50.3	18.2

2. Introduction

2.1 Background

In January 2009, the Nebraska Department of Education (NDE) contracted with Data Recognition Corporation (DRC) to provide and operate a computerized information system to support the administration, record keeping, and reporting for statewide student assessment and accountability under the direction of the Department of Education.

NeSA Content Areas and Grade Levels: Legislative Bill (LB) 1157 passed by the 2008 Nebraska Legislature (http://uniweb.legislature.ne.gov/FloorDocs/Current/PDF/Slip/LB1157.pdf) requires a single statewide assessment of the Nebraska academic content standards for writing, reading, mathematics, and science in Nebraska's K-12 public schools. The new assessment system is named NeSA (Nebraska State Accountability) with NeSA-R for reading assessments. Reading assessments were administered in grades 3 through 8 and 11 for the first time in the spring of 2010.

Phase-In Schedule for NeSA: The NDE prescribed such assessments starting in the 2009-2010 school year to be phased in as shown in Table 2.1.1. The state used the expertise and experience of in-state educators to participate, to the maximum extent possible, in the design and development of the new statewide assessment system. NDE developed the NeSA-R tests for use in the state accountability system and was charged with setting student academic performance level standards on the NeSA-R tests.

Table 2.1.1: NeSA Administration Schedule

Content Area	Administration Year		Grades
	Field Test	Operational	
Reading	2009	2010	3 through 8 and one high school
Mathematics	2010	2011	3 through 8 and one high school
Science	2011	2012	Elementary, middle/junior high, high school

NDE required standard-setting procedures to determine student academic performance levels for the NeSA-R assessments administered to each of grades 3 through 8 and 11. DRC, with the assistance of NDE, organized and facilitated the standard setting events.

For NeSA-R, there are three student performance levels: Below the Standards, Meets the Standards, Exceeds the Standards, therefore establishing two cut points. For federal reporting purposes, *Proficiency* is defined as students performing at Meets the Standards and Exceeds the Standards levels. These labels were chosen by the State Board of Education (SBE) after the standard setting events; the labels used during the events were Basic, Proficient, and Advanced.

2.2 Purpose and Objectives of NeSA and Standard Setting Event

NeSA-R tests will assess the State-adopted academic standards to promote student learning and to measure student performance on state academic standards, as well as to:

1. identify areas in which students, schools, or school districts need additional support;

2. indicate the academic achievement for schools, districts, and the State;
3. satisfy federal reporting requirements; and
4. provide professional development to educators.

The results from the NeSA-R tests will be used for evaluating Adequate Yearly Progress (AYP) for No Child Left Behind (NCLB) and for reporting annual State school and district ratings of end-of-year performance.

The panelists who participated in the standard setting were reminded of the role of NeSA at the start of the process. They were further told that their role was to develop a recommendation on the performance standards that would be presented to the SBE for consideration and possible adoption.

There are a multitude of standard setting methods that have been proposed over the decades. These fall into two major approaches:

1. *Item-based*, which focus on what knowledge, skills, and behaviors are required to successfully respond, and
2. *Student-based*, which focus on what proficiencies individual students possess.

For the NeSA, both approaches were used to set the standards. The method of record was the item-based Bookmark method. A Contrasting Groups survey of Nebraska teachers was also used to validate and strengthen the Bookmark results.

2.3 Bookmark Standard Setting Method

DRC utilized a Bookmark procedure, following closely the method suggested by Lewis, Mitzel, and Green (1996). Bookmark is one in a broad category of methods commonly referred to as item mapping, which focuses on items rather than examinees. The essential task is to identify the items that can be answered successfully (67% likelihood) by students at the boundaries of the performance levels. The logit difficulty value that separates the items that students can do from those they cannot do establishes the bookmark cut score.

All panelists were trained in a large group prior to breaking into smaller working groups. Training covered the following points:

- The bookmark represents a judgment of the divide between items that a student at the threshold of a performance level should master from those it is not necessary to master.
- Bookmark placement should not be thought of as separating two items, but rather two groups of items. In other words, a placement should not hinge on distinctions drawn for adjacent items, without some compelling reason, such as a large gap in content difficulty.
- Students at a given cut score will have approximately a 0.67 probability of correctly responding to a multiple-choice item also at the cut score. These same students will have a higher probability of success on easier items (before the bookmark placement) and a lower probability of success on harder items (after the bookmark placement).
- In placing their bookmarks, the task was to consider what students *should* know and be able to do in the context of the skills implied by the Performance Level Descriptors and the item content.

- Panelists were instructed to start with placing the Below the Standards/Meets the Standards boundary and then the Meets the Standards/Exceeds the Standards boundary.
- Panelists were asked to record their bookmark placements on the rating form. The judgments were entered into a spreadsheet program, and the median cut score was calculated for the full panel.

To begin the process, participants were asked to visualize the knowledge and skills of a student who is at the borderline between two Performance Levels based on the performance level descriptors (PLD's). Participants were given a booklet with items ordered from least to most difficult. In addition, panelists were also provided with supporting materials for each item including the correct response, content objective, and item sequence in the test booklets.

The task for the panelist was to proceed through the ordered item booklet (OIB) and ask whether the borderline student could answer each item. Each panelist placed a bookmark at the page in the booklet where they felt the borderline student had not mastered the item. *Mastery* was defined as having at least a 67% likelihood of responding correctly.

The DRC adaptation of the Bookmark procedure involved three rounds of deliberation, discussion, and feedback. These iterations are described in more detail in Section 4.

2.4 Contrasting Groups Standard Setting Method

An examinee-based Contrasting Groups (Cizek & Bunch, 2007) survey was included to compliment the item-based Bookmark method. All Nebraska reading teachers and specialists were invited to participate in the survey, which asked them to evaluate each student with whom they were familiar and indicate which performance level best described the student. The survey was conducted prior to the first operational administration of the NeSA-R, so ratings were determined by the teachers' firsthand experience with the students in the classroom, not their performance on the test.

The survey was available online and teachers had the opportunity to select students from their own school and to exclude any students they chose. The instructions emphasized the importance of knowing the student and the student's status. Teachers were encouraged to omit ratings for any student for whom the teacher did not have firsthand knowledge.

The results of the survey were summarized, shared with the Bookmark panels, and presented to SBE with the final cut score recommendations.

2.5 Meeting with a Committee of Stakeholders

In preparation for the July 8, 2010 Board meeting, DRC presented to a subgroup of Board Committee members, media and other stakeholders on February 25, 2010. The purpose of the July meeting was to formally adopt an anticipated motion establishing cut scores for the NeSA-R based on results from the two standard setting events and on recommendations from the NDE. In contrast, the February meeting was a preview of the July meeting. This meeting allowed the participants to familiarize themselves with the standard setting process prior to introducing standard setting results. This involved DRC presenting an overview of the standard setting processes and the appropriate interpretation of the results from the studies. In addition, there was a discussion of the information needed and effective methods for its interpretation to make a sound policy decision.

3. Preparation for Standard Setting

In April 2010, a standard setting plan was proposed by DRC. The plan was reviewed and approved by NDE and its Technical Advisory Committee (TAC). The plan described the purpose of the meeting, specifications of panelists, methodology, and potential consequences related to accountability. This section provides an overview of relevant sections from the plan.

3.1 Bookmark Panelist Recruitment

NDE recruited panelists for the Standard Setting process through a series of steps.

- In January of 2010, Dr. Pat Roschewski communicated with District Assessment Contacts, informing them of the plan for establishing NeSA-R cut scores and the need for Nebraska educators to participate in the process. Additionally, information regarding the Standard Setting process was communicated to Nebraska districts in Standards, Assessment, and Accountability Updates.
- The Statewide Assessment Office posted an application for participation in the Standard Setting process on its website. Individuals interested in participating completed the application and submitted it by March 15, 2010.
- A committee comprised of Statewide Assessment team members determined participants through a review of all applications received. Three criteria were considered:
    1. Educational role.
    2. Geographic location.
    3. Knowledge of and experience with administration of NeSA-R.
- Applicants received communication from the Statewide Assessment Office by April 1, 2010, informing them of their selection status.

A total of 101 panelists participated in the Bookmark event. Table 3.1.1 summarizes information about characteristics of the participating panelists based on their self-reported responses to the Participant Survey. Most panelists were classroom teachers. A few were non-teacher educators. While the group was predominantly female, this reflects the reality of reading instruction.

Table 3.1.1 Panelist Summary

Demographic		Reading
Gender	**Male**	14
	Female	87
Ethnicity	**White/non-Hispanic**	98
	Multi-racial/ethnic	2
	Latino/Hispanic	1
Role	**Other**	5
	Teacher	83
	Educator	13
Region	**Rural**	60
	Urban	21
	Suburban	13
Experience	**0 - 5 years**	15
	6 - 10 years	18
	11 - 15 years	17
	16 – 20 years	17
	21 – 25 years	13
	26 – 30 years	9
	31 – 35 years	7
	> 36 years	5
Total *N*		101

3.2 Roles and Responsibilities

A successful standard setting requires the concerted and coordinated efforts of many people including staff from NDE and DRC, and, most importantly, the panelists. Roles and responsibilities are briefly summarized below:

Panelists—brought their unique and individual educational experience and expertise to develop recommendations for defining the performance levels for the NeSA-R by applying the procedures as directed by the room facilitators. Their knowledge of reading instruction and curriculum in Nebraska and their familiarity with Nebraska students were the basis for the validity of the recommended performance standards.

Nebraska Department of Education (NDE)—convened the meeting and introduced the NeSA-R program and the importance of standard setting. NDE staff monitored the progress of each panel and fielded questions on the assessment and test content and, more generally, on any policy concerns.

DRC Staff—facilitated the sessions and provided logistical and technical support.

Psychometric Lead—introduced procedures during training and monitored progress and results during the event.

Room Facilitators—reviewed procedures, kept panels moving at a pace that would achieve agenda timelines, and explained results.

Test Development Specialists—assisted as needed with the Performance Levels and covered questions about test content.

Data Analyst—captured the panelists' bookmark settings and performed the necessary psychometric analyses.

Project Management—maintained security of materials through check-in and check-out procedures, liaison with hotel facility staff, and overall coordination of meeting logistics.

3.3 Materials Preparation

Workshop materials were developed and printed by DRC. Following is a list of materials made available to panelists during the workshop:

- Training Materials
- Operational Test Forms
- Ordered Item Booklet (OIB)
- Performance Standards
- Item Map
- Item Separation Map
- Participant Rating Forms
- Stationery Supplies.

Training materials, including the sample ordered item booklet, item map, item separation map, and rating form were developed and printed by DRC staff. The training materials were developed using items and item data from the NAEP website.

Reading Performance Level Descriptors were originally developed by the NDE with assistance from educators in the field. Please see Appendix A for a complete listing of the PLD's.

3.4 Ordered Item Booklet Item Placements

The task presented to the panelists was to identify the item in the Ordered Item Booklet for which the student on the boundary between two levels can no longer answer the item correctly. The required level of mastery was defined operationally as a probability of success of 0.67. With the Rasch model, the choice of the mastery level does not affect the ordering of the items, but it does affect which Scale Score aligns with the bookmarked item.

The Rasch model for dichotomous items (Wright & Stone, 1979) defines the probability of success as:

1. $p = \frac{e^{b-d}}{1+e^{b-d}}$.

With a little algebra when p = 0.67, this implies the logit cut score is shifted by *0.69 logits* from the logit difficulty of the bookmarked item:

2. $(b-d) = ln\frac{0.67}{0.33} = ln(2) = 0.69$.

3.5 Ordered Item Booklet Preparation

Each Ordered Item Booklet (OIB) contained all items in the grade in order of item difficulty from least to most difficult, based on item difficulties obtained from the spring 2010 NeSA-R administration. Table 3.5.1 displays the number of items/score points per grade on the operational forms. Item Separation Charts for each grade are included in Appendix E.

Table 3.5.1: Number of Score Points in Ordered Item Booklet

Content	Grade	No. of Score Points in the OIB
Reading and Research	**3**	45
	4	45
	5	48
	6	48
	7	48
	8	50
	11	50

4. Standard Setting Procedures

4.1 Contrasting Groups Procedure

An examinee-based Contrasting Groups survey was included to complement the item-based Bookmark method. All Nebraska reading teachers were invited to participate in the survey, which was presented online. The task for the teachers was to evaluate each student with whom the teacher was familiar and indicate the performance level that best described the student. The survey was conducted prior to the first operational administration of the NeSA-R, so ratings were determined by the teachers' firsthand experience with the students in the classroom, not their performance on the test. At the time the survey was done, the performance level labels being used were Advanced, Proficient, and Basic. A draft of the performance level descriptors (PLD's) was available online for review at any point in the process.

The teachers had the opportunity to select students from their own classes and schools and to exclude any students they chose. The instructions emphasized the importance of knowing the student and the student's status. Teachers were encouraged to omit ratings for any students for whom they did not have firsthand knowledge.

Recruitment: In December 2009, NDE and DRC contacted Nebraska District Assessment Coordinators (DAC's) to solicit their cooperation in the study that would bring teachers' knowledge of reading instruction and an understanding of their students together. The DAC's were first asked to provide contacts for these reading teachers and specialists.

In early February 2010, DRC sent an initial invitation to teachers. This invitation asked for their participation in an online study that would use their professional judgment to help establish the performance levels for the NeSA-R. The teachers were assured that they would be provided training via WebEx prior to participating, that it should take less than 30 minutes of their time, and that their responses were confidential. They were also given the schedule for the survey and the training sessions.

A follow-up email was sent to the participating teachers at the end of February reminding them of the WebEx dates, sign-on and times, and information on the online delivery system, eDIRECT.

Training: DRC hosted ten WebEx sessions to introduce teachers to the online contrasting group survey. For teachers who were unable to attend a WebEx session, NDE placed the training materials on its Website on March 17, 2010. The WebEx sessions were interactive, allowing teachers to pose questions and seek immediate clarification. Typically, the sessions lasted fifteen to twenty minutes. Feedback on the training was positive, but there were requests for scheduled times more convenient for the Mountain Time Zone.

The training covered the details of navigating the survey website, saving the work, returning after interruptions, and submitting the ratings. In the training sessions and in the online instructions, each teacher was asked to:

- Use the school and district rosters provided to create a personal class roster with 25-30 students representing all performance levels.
- Note the instructions at the top of each page of the survey.
- Read and refer back to the performance level descriptors in the course of the survey.

- Complete the survey as soon as possible after training, but no later than March 26, 2010.

Table 4.1.1: WebEx Training Schedule

SESSION	DATE	TIME
1	Wednesday, March 10, 2010	7:00 – 7:30 AM
2	Wednesday, March 10, 2010	3:30 – 4:00 PM
3	Thursday, March 11, 2010	9:00 – 9:30 AM
4	Thursday, March 11, 2010	4:00 – 4:30 PM
5	Friday, March 12, 2010	11:00 – 11:30 AM
6	Friday, March 12, 2010	1:00 – 1:30 PM
7	Monday, March 15, 2010	7:00 – 7:30 AM
8	Monday, March 15, 2010	2:30 – 3:00 PM
9	Tuesday, March 16, 2010	3:00 – 3:30 PM
10	Tuesday, March 16, 2010	4:00 – 4:30 PM

The instructions explicitly informed teachers that they were not required to select students with whom they had little experience nor did they need to rate students, even if selected, if they were uncomfortable assigning the student to a performance level for any reason.

Survey Results: Appendix F provides detailed summaries of the teacher survey, including breakouts by gender, ethnic group, English language learners (ELL), and free lunch status (FLS). The tables also show the agreement between the teacher ratings and the performance level assignments using the final, SBE-approved cut scores. The correlations were about 0.6 or higher across the grades. It is worth reiterating that the survey was conducted prior to the first operational assessment, while the PLD's were in draft form, and there was no facilitated group discussion of the PLD's.

A total of 413 teachers participated in the survey. The distribution across grades was acceptable but lower than targeted, ranging between a high of 81 for grade 3 and a low of 42 for grade 7. The initial target number was 100 per grade. Recruiting strategies are being reviewed to obtain higher participation in 2011. Feedback from the participants indicated the task was easier and took less time than they expected. The breakdown by grade is given in Table 4.1.2.

Table 4.1.2: Contrasting Groups Participation by Grade

Grade	Number of Teachers	Number of Students Rated
3	81	1424
4	71	1437
5	64	1096
6	54	1200
7	42	991
8	50	1262
11	51	1407
Total	**413**	**8817**

The cut scores were derived as the point on the scale score metric where the higher performance level became more likely than the lower level for students with the same estimated abilities. The likelihood for "Below the Standards" is shown in Table 4.1.3 as the ratio of the number in the Below group divided by the total numbers in Meets and Exceeds. There is some ambiguity about the exact logit value of the cut score

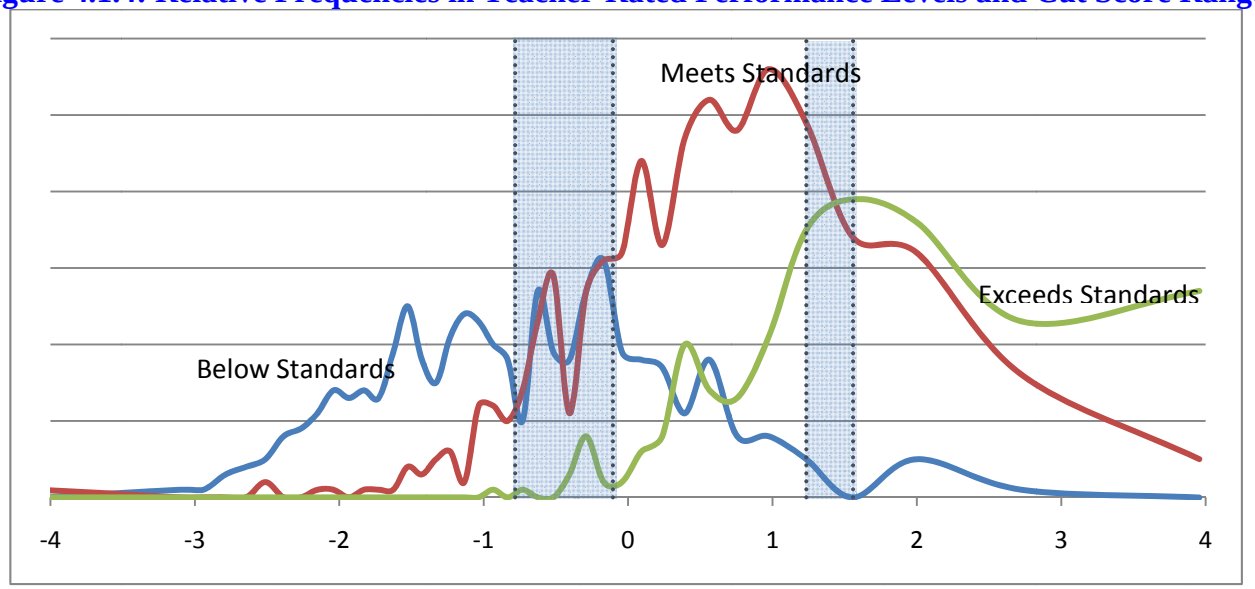
because there exact point will fall between two raw scores and because they will typically be some fluctuation in the observed counts. This is illustrated for grade 3 in Table 4.1.3 and graphically in Figure 4.1.4.

Table 4.1.3: Calculation of Grade 3 NeSA-R Contrasting Groups Cut Scores

Raw Score	Number Below	Number Meets	Number Exceeds	Likelihood Below Std	Likelihood Meets Std	Logit Ability
25	23	12	0	0.66	1.00	-1.034
26	20	12	1	0.61	0.92	-0.935
27	18	10	0	0.64	1.00	-0.833
28	10	14	1	0.40	0.93	-0.730
29	27	23	0	0.54	1.00	-0.625
30	19	29	0	0.40	1.00	-0.517
31	18	11	3	0.56	0.79	-0.405
32	27	27	8	0.44	0.77	-0.290
33	31	31	2	0.48	0.94	-0.169
34	19	32	2	0.36	0.94	-0.042
35	18	44	6	0.26	0.88	0.093
36	17	33	8	0.29	0.80	0.237
37	11	47	20	0.14	0.70	0.393
38	18	52	14	0.21	0.79	0.564
39	8	48	13	0.12	0.79	0.756
40	8	56	21	0.09	0.73	0.975
41	5	49	35	0.06	0.58	1.234
42	0	34	39	0.00	0.47	1.558
43	5	32	36	0.07	0.47	1.999
44	1	16	23	0.03	0.41	2.727
45	0	5	27	0.00	0.16	3.956

The likelihood of level Meets the Standards becomes less likely than level Exceeds the Standards between raw scores 41 and 42, which correspond to logits of 1.234 and 1.558. Any logit value in this range would be consistent with the teacher ratings. The change between Below and Meets is even less certain because the likelihood in this range doesn't decrease smoothly. Any cut score between raw score 28 and 33 (logits -0.73 to -0.17) might be argued, although typically some form of smoothing and interpolation will be applied.

Figure 4.1.4: Relative Frequencies in Teacher-Rated Performance Levels and Cut Score Ranges



4.2 Modified Bookmark Procedure

The agenda for the bookmark event is presented in Appendix B.1. The process, including training, was completed in three days, Monday through Wednesday, June 28-30, 2010, using three grade-grouped panels: lower, middle, and high school. The intent of the grade groupings was to ensure panelists worked with content with which they were familiar while giving each panel more breadth, and the result more continuity across grades. The precise groupings were realigned during the event to best match panelists to their grade. The groupings and timing are diagramed in Appendix B.2.

Training was conducted Monday morning with a single trainer for a single large group of the three panels. A copy of the PowerPoint slides used for training is presented in Appendix C. Training materials included:

- Performance Level Descriptors (PLD's)
- Ordered Item Booklets (OIB)
- Item Map
- Item Separation Chart
- Rating Form

Participants were told that:

- all materials were secure and were not to leave the meeting room,
- the bookmark placement should reflect the panelist's own opinion and not the group consensus, and
- they should contribute their own personal experience and expertise to better inform the group discussion and recommendation; consensus was not necessary.

The critical objective of the training was to ensure the panelists understood the task being presented to them. Components included an overview of their role in the process, a detailed description of all steps in the Bookmark method, and a practice exercise based on a short test form drawn from NAEP materials. The point of the practice exercise was to provide hands-on experience with the steps and allow the panelists to receive any additional explanation they needed or requested.

Panelists were told that the process would include three iterations (rounds) of individual judgments, large group discussions between rounds, and opportunities to revise individual judgments. After the second and third rounds, panelists would have the opportunity to review impacts in the form of percent of students in each performance level, resulting from the group recommendation. In addition, panels for the appropriate grades would be shown relevant NAEP and ACT statistics.

After the training and practice exercise, the panelist broke into the smaller groups and began work on specific grades. The process began with a review of the PLD's specific to that grade to sharpen the understanding of what was expected of students at each level. The panelists then worked through the spring operational form of NeSA-R. This task was included to give panelists a direct appreciation of the students NeSA-R experience. They were encouraged to take notes concerning their impressions of the items. After a short discussion of the operational form, the actual bookmarking began.

Round 1. Round 1 began after the review of items and passages. Participants reviewed the ordered item booklets independently to ensure the initial bookmarks were independent of other panelists' opinions. During this review, they were asked to determine the knowledge, skills, and competencies required to respond correctly to each progressively more difficult item and when these requirements exceed the capabilities of Below the Standards, Meets the Standards, or Exceeds the Standards level students. It was emphasized that the work for this round was to be individual.

The panelists were reminded periodically that the bookmarks are placed so that the borderline student has mastered those before the bookmark and not those after the bookmark. To reduce counter-productive argument about the placement of specific items in the OIB, panelists were informed that the placement was empirical based on the spring assessment and that they should focus on ranges of items rather than the details of individual items.

Round 2. The results from Round 1 were presented and explained at the beginning of Round 2. The bookmark page numbers for each panelist, the median page number of the full panel, the distribution of cut scores for each performance level, and the *impact* data were presented to the panelists. The impact data was simply the percentage of students placed in each performance level based on Spring 2010 NeSA-R student performance and Round 1 panelists' recommendations. Panelists were then asked to provide rationales for their Round 1 placements and what skills and knowledge were required. During the discussion, there was no attempt to achieve consensus; the bookmark placements were to reflect the opinions of the individual panelists.

After the group discussion, panelists were given the opportunity to revise their bookmark placements. The placements were again collected and used to calculate revised cut scores and impact data for the full panel.

Round 3. Round 3 began with the presentation of Round 2 results and the relevant contrasting groups data. When applicable to grade, the NAEP (grades 4 and 8) and ACT (grade 11) data were also provided. Again, panelists were instructed to explain the thinking for their Round 2 placements in terms of the skills and knowledge required. Following the discussion, the panelists made any final adjustment to their individual placements. These ratings were recorded and used to produce the final group recommendation.

4.3 Vertical Articulation Across Grades

For accountability and monitoring longitudinal progress, it is important that the performance levels are coherent across grades. One would expect, for example, that the percent meeting or exceeding the standards would be *consistent*, perhaps trending up or down but not fluctuating erratically. This becomes more critical when performance levels with high stakes consequences are established for contiguous grades.

Three distinct tactics were used to achieve a satisfactory degree of coherence. First, the common introduction and training for all panelists ensured a common understanding of the PLD's and the bookmarking task. Second, the grade groupings ensured the panelists were familiar with, and participated in, the deliberations and recommendations for adjacent grades. This was enhanced by large group sessions each morning that allowed for more general, cross-grade discussion. Finally, after the panelists completed their work, the group recommendations were statistically smoothed to achieve coherent percents in each performance level. This approach allowed the data from all grades to be considered simultaneously. Any

trend over grades was established by the panels, but it was assumed that the entire body of data was more reliable than any one grade.

As a practical matter, no adjustment to a grade was allowed that was greater than one standard error, and the sum of the adjustments across grades was restricted to one tenth of a standard error. The final cut score recommendations were obtained by interpolating the logit cut scores to obtain the target percentages.

4.4 Merging Bookmark and Contrasting Groups

The item-based Bookmark method was the designated method of record. The Bookmark results were the crux of the recommendation to the SBE. The recommendation was developed by experts on education in Nebraska, primarily classroom teachers, from their understanding of the PLD's, and their assessment of the knowledge, skills, and behaviors required by the operational items.

The Contrasting Groups survey involved a different sample from the same population of experts. The focus for this method was on students known to the teacher and on the performance level best describing each of those students, independent of any assessment. While the PLD's were available on demand as a pop-up for the participants in the Contrasting Groups, there was no group training to ensure a common understanding of the PLD's. However, the data are too rich to be ignored.

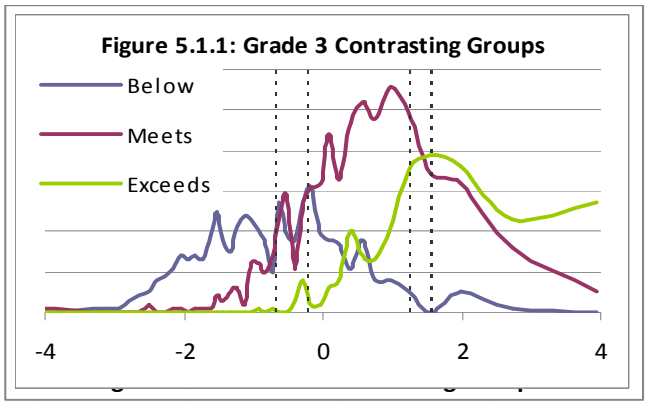
The final recommendation to the SBE was based on a composite that used both sets of data with smoothing. Details of the arithmetic are included in the results section, but the recommended cut scores did not differ from the Bookmark result by as much as one standard error.

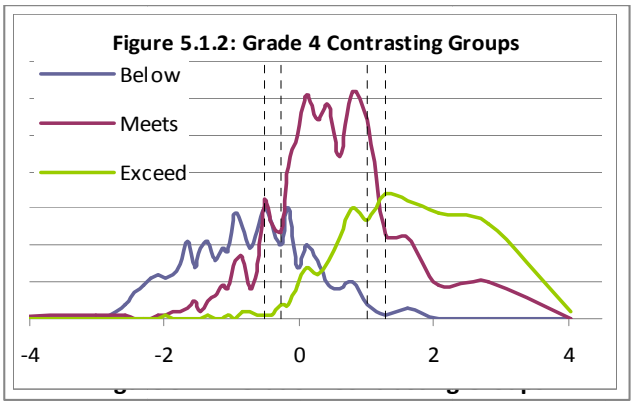
5. Results

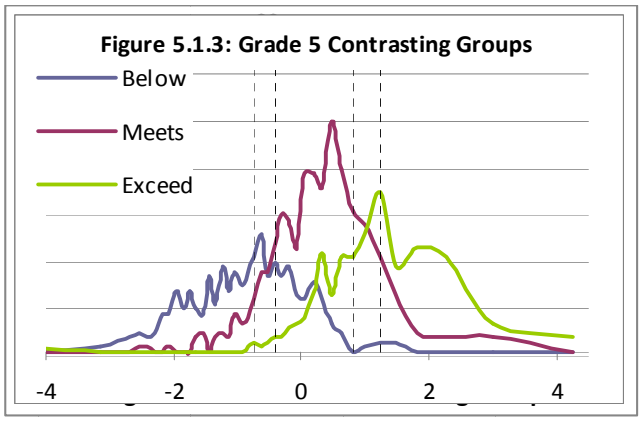
5.1 Contrasting Groups Analyses

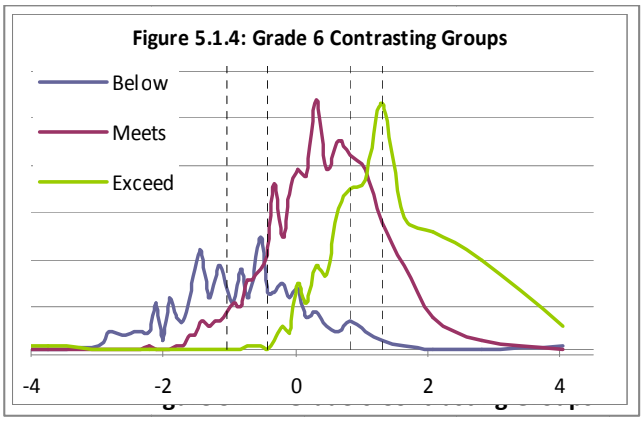
The estimated cut scores were derived from the Contrasting Groups survey by locating the point on the scale for which the likelihood of being in the higher performance level surpasses the likelihood of being in the lower level. For the cut between Meets the Standards and Exceeds the Standards, this is accomplished by locating the number correct score x for which more students who scored x were rated in the Exceeds category than were rated in the Meets category. Tables summarizing the distributions of the ratings are presented in Appendix G. Just a reminder that the panelists were provided different level names at the time of the survey: Basic, Proficient and Advanced.

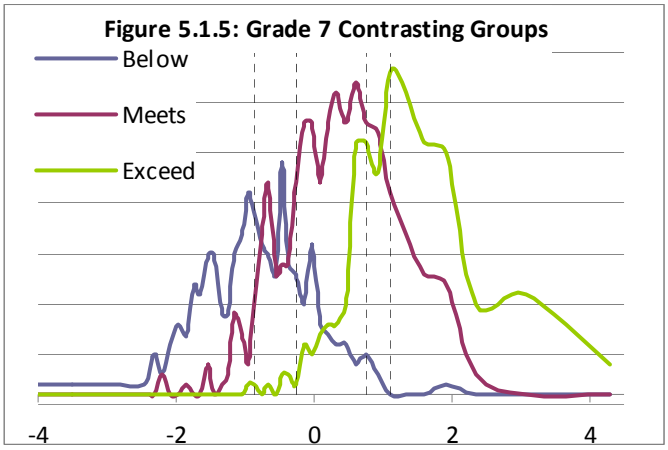
The same data are presented graphically below.

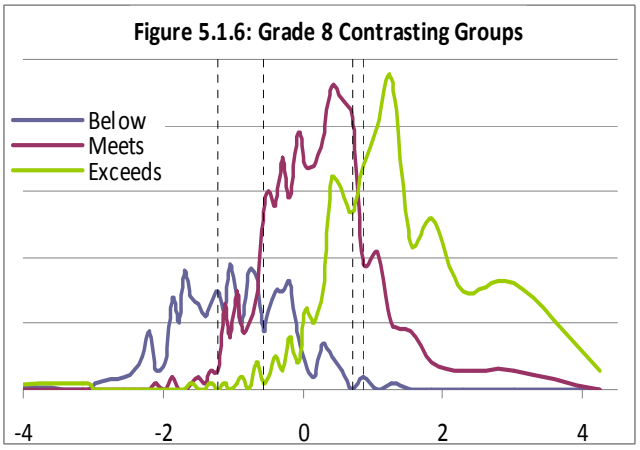


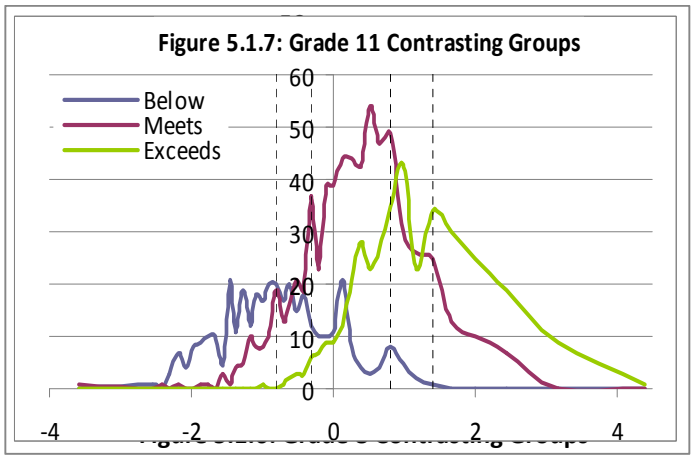












5.2 Bookmark Analyses

The bookmark pages, determined by the 40 to 60 panelists, formed the crux of the recommended Scale Score cut points. The bookmarks from the panelists were summarized using medians to minimize the effect of extreme values. The medians and their standard errors are shown below in Table 5.2.1.

Table 5.2.1: Bookmark Page Number Medians and Standard Errors

	Number of Panelists	Rd 1 B/M	Rd 1 M/E	Rd 2 B/M	Rd 2 M/E	Rd 3 B/M	Rd 3 M/E
Grade 3	41						
Median		15	36	15	37	15	41
Std Dev		3.74	3.81	2.78	2.69	2.36	2.66
SE (med)		0.73	0.74	0.54	0.52	0.46	0.52
Grade 4	41						
Median		12	34	11	35	11	39
Std Dev		4.22	4.48	2.45	2.93	3.76	2.96
SE (med)		0.82	0.88	0.48	0.57	0.73	0.58
Grade 5	41						
Median		14	41	14	41	14	41
Std Dev		3.30	4.30	2.40	3.00	1.60	2.90
SE (med)		0.60	0.80	0.50	0.60	0.30	0.60
Grade 6	33						
Median		13	41	15	44	16	44
Std Dev		4.50	4.70	3.50	3.20	3.30	3.30
SE (med)		1.00	1.00	0.80	0.70	0.70	0.70
Grade 7	33						
Median		14	38	12	38	14	40
Std Dev		4.14	4.11	1.05	0.68	1.21	2.37
SE (med)		0.90	0.89	0.23	0.15	0.26	0.52
Grade 8	61						
Median		17	42	17	44	17	44
Std Dev		4.79	4.49	3.15	3.07	3.19	2.75
SE (med)		0.77	0.72	0.50	0.49	0.51	0.44
Grade 11	27						
Median		19	36.5	19.5	38	20	42
Std Dev		5.44	4.04	3.30	2.95	4.33	1.91
SE (med)		1.29	0.95	0.78	0.70	1.02	0.45

Each bookmark page number is an item location, which implies a logit difficulty value. The logit difficulties determine the raw score and scale score cut points. The scale score cut and its standard error of measurement (SEM) were used to establish the *1 SEM* confidence intervals around the recommended cut score. NDE used the standard errors to identify the appropriate cut score taking into consideration variance in the human judgments and imprecision in the test itself.

5.3 Recommendation and Approval of State Board of Education

The State Board of Education (SBE) reviewed the results from both the Bookmark and Contrasting Groups studies. While the SBE was initially more comfortable with the results from the Contrasting Groups study in terms of the outcomes, DRC presented the third option of a simple, unweighted averaging of the logit cuts from the two studies. The average was computed in the logit metric and translated into percent of students in category. The percent in categories was *not* the statistic of focus; these were calculated after the logit cuts were determined.

Two notable adjustments were made to the third option to arrive at the final cut scores:

1) grade 8 was adjusted in "Exceeds the Standards" from 27.4 percent to 22.2 percent to more closely match the other grades, and,
2) all grades except grade 7 were adjusted to allow more Below the Standards students in the category and correspondingly fewer students in the Meets the Standards category.

Summary values for the cut scores and impacts are shown in Table 5.3.1 with details presented in Appendix H.

Table 5.3.1: Logit and 2010 Raw Score Cut points for NeSA-R

	Logit Cut points		2010 Raw Score Ranges by Performance Level			Percent in Each Performance Level		
Grade	B/M	M/E	Below	Meets	Exceeds	Below	Meets	Exceeds
3	-0.5168	1.2340	0 to 29	30 to 40	41 to 45	32.5	47.4	20.1
4	-0.5117	0.8591	0 to 29	30 to 39	40 to 45	30.5	48.1	21.4
5	-0.4122	0.8560	0 to 31	32 to 41	42 to 48	32.6	48.2	19.2
6	-0.4331	0.8924	0 to 32	33 to 42	43 to 48	31.8	48.6	19.6
7	-0.5104	0.7855	0 to 29	30 to 40	41 to 48	31.0	48.0	21.0
8	-0.4812	0.8712	0 to 32	32 to 42	43 to 50	29.6	48.1	22.2
11	-0.4103	0.8508	0 to 31	32 to 42	43 to 50	31.5	50.3	18.2

The Scale Score metric was derived from the logits so that the minimum Scale Score for Meets the Standards was 85 and the minimum score for Exceeds the Standards was 135 for all grades. It is anticipated that the 85 and 135 values will be maintained for the remaining content areas as well. The calculations for the NeSA-R Scale Score conversion are in Table 5.3.2.

Table 5.3.2: Conversion of Logits to Scale Scores

	Logit Cutpoints		Scale Score Ranges by Performance Level			Conversion	
Grade	B/M	M/E	Below	Meets	Exceeds	Slope	Intercept
3	-0.5168	1.2340	1 to 84	85-134	135 to 200	28.55837	99.25997
4	-0.5117	0.8591	1 to 84	85-134	135 to 200	36.47505	103.16528
5	-0.4122	0.8560	1 to 84	85-134	135 to 200	39.42751	100.75302
6	-0.4331	0.8924	1 to 84	85-134	135 to 200	37.72161	100.83823
7	-0.5104	0.7855	1 to 84	85-134	135 to 200	38.58471	104.19271
8	-0.4812	0.8712	1 to 84	85-134	135 to 200	36.97131	102.29159
11	-0.4103	0.8508	1 to 84	85-134	135 to 200	39.64793	100.76854

5.4 Panelists' Survey Evaluation Results

On the last day of the standard setting, panelists were asked to complete an evaluation on the standard setting meeting itself. This information was used to assess the panelists' impression of the validity of the process and their confidence in the result. A copy of the instrument is included in Appendix I and a summary of the results is included Appendix J.

6. References

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.

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

Wright, B. & Stone, M. (1979). *Best test design*. Chicago: MESA Press.

Appendices

Appendix A: NeSA-R Performance Level Descriptors

The Performance Level Descriptors (PLD's) provide meaning to the Scale Score metric and give a qualitative description of the numeric scores. The attached PLD were used by the panelists both during the standard setting Bookmark and the contrasting groups studies. The labels used for the levels were *Basic*, *Proficient*, and *Advanced* at the time of standard setting. They were changed before reporting to *Below the Standards*, *Meets the Standards*, and *Exceeds the Standards.*

Grade 3

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor Grade 3
Advanced Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above third grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level. An *advanced* learner: • Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text. • Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade level vocabulary. • Has a thorough understanding of author's purpose. • Consistently recognizes how story elements (e.g., plot, setting, characterization, problems) impact text. • Consistently distinguishes stated or implied main idea and relevant details in informational text. • Consistently identifies and uses literary devices (e.g., simile, alliteration, onomatopoeia, rhythm). • Consistently identifies and uses organizational patterns of informational text (e.g., sequence, description, cause/effect, compare/contrast). • Consistently interprets informational text features (e.g., headings, maps, timelines). • Consistently identifies defining characteristics of narrative and informational genres (e.g., poetry, biographies, historical fiction). • Consistently answers literal and inferential questions with accuracy and provides supporting information.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 3

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at third grade. A student scoring at the proficient level *generally* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a sufficient understanding of author's purpose.
- Generally recognizes how story elements (e.g., plot, setting, characterization, problems) impact text.
- Generally distinguishes stated or implied main idea and relevant details in informational text.
- Generally identifies and uses literary devices (e.g., simile, alliteration, onomatopoeia, rhythm).
- Generally identifies and uses organizational patterns of informational text (e.g., sequence, description, cause/effect, compare/contrast).
- Generally interprets informational text features (e.g., headings, maps, timelines).
- Generally identifies defining characteristics of narrative and informational genres (e.g., poetry, biographies, historical fiction).
- Generally answers literal and inferential questions with accuracy.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 3

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at third grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade level vocabulary.
- Has an insufficient understanding of author's purpose.
- Inconsistently recognizes how story elements (e.g., plot, setting, characterization, problems) impact text.
- Inconsistently distinguishes stated main idea and some details in informational text.
- Inconsistently identifies and uses literary devices (e.g., simile, alliteration, onomatopoeia, rhythm).
- Inconsistently identifies organizational patterns of informational text (e.g., sequence, description, cause/effect, compare/contrast).
- Inconsistently interprets informational text features (e.g., headings, maps, timelines).
- Insufficiently identifies defining characteristics of narrative and informational genres (e.g., poetry, biographies, historical fiction).
- Inconsistently answers literal questions with accuracy.

Grade 4

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 4

Advanced

Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above fourth grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level.

An *advanced* learner:

- Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text.
- Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a thorough understanding of how an author's purpose and perspective (beliefs, assumptions, biases) influence text.
- Consistently recognizes and analyzes how story elements (e.g., plot, setting, characterization, problem/resolution) impact text.
- Consistently determines stated or implied main idea and relevant details in informational text.
- Consistently identifies and uses literary devices (e.g., simile, alliteration, metaphor).
- Consistently identifies and uses organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Consistently interprets informational text features (e.g., headings, maps, tables).
- Consistently identifies defining characteristics of narrative and informational genres (e.g., poetry, biographies, folk tales).
- Consistently answers literal, inferential, and critical questions with accuracy and provides supporting information.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 4

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at fourth grade. A student scoring at the proficient level *generally* utilizes a variety of reading strategies to comprehend and interpret grade-level appropriate narrative and informational text.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar words.
- Has a sufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) influence text.
- Generally recognizes and analyzes how story elements (e.g., plot, setting, characterization, problem/solution) impact text.
- Generally determines stated or implied main idea and relevant details in informational text.
- Generally identifies and uses literary devices (e.g., simile, alliteration, metaphor).
- Generally identifies and uses organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Generally interprets informational text features (e.g., headings, maps, tables).
- Generally identifies defining characteristics of narrative and informational genres (e.g., poetry, biographies, folk tales).
- Generally answers literal, inferential, and critical questions with accuracy.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 4

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at fourth grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has an insufficient understanding of how an author's purpose influences text.
- Inconsistently recognizes how story elements (e.g., plot setting, characterization, problem/solution) impact text.
- Inconsistently distinguishes stated main idea and relevant details in informational text.
- Inconsistently identifies and uses literary devices (e.g., simile, alliteration, metaphor).
- Inconsistently identifies and uses organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Inconsistently interprets informational text features (e.g., headings, maps, tables).
- Inconsistently identifies defining characteristics of narrative and informational genres (e.g., poetry, biographies, folk tales).
- Inconsistently answers literal and inferential questions with accuracy.

Grade 5

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 5

Advanced

Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above fifth grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level.

An *advanced* learner:

- Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text.
- Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade level vocabulary.
- Has a thorough understanding of how an author's purpose and perspective (beliefs, assumptions, biases) influence text.
- Consistently recognizes and analyzes how story elements (e.g., plot, setting, characterization, theme) impact text.
- Consistently summarizes and analyzes stated or implied main idea and relevant details in informational text.
- Consistently identifies and uses literary devices (e.g., simile, alliteration, metaphor, imagery).
- Consistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Consistently interprets informational text features (e.g., headings, maps, indexes).
- Consistently identifies defining characteristics of narrative and informational genres (e.g., poetry, myths, fantasies).
- Consistently answers literal, inferential, critical, and interpretive questions with accuracy and provides supporting information.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 5

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at fifth grade. A student scoring at the proficient level *generally* utilizes a variety of reading strategies to comprehend and interpret grade-level appropriate narrative and informational text.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a sufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) influence text.
- Generally recognizes and analyzes how story elements (e.g., plot, setting, characterization, theme) impact text.
- Generally summarizes and analyzes stated or implied main idea and relevant details in informational text.
- Generally identifies and uses literary devices (e.g., simile, alliteration, metaphor, imagery).
- Generally applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Generally interprets informational text features (e.g., headings, maps, indexes).
- Generally identifies defining characteristics of narrative and informational genres (e.g., poetry, myths, fantasies).
- Generally answers literal, inferential, critical, and interpretive questions with accuracy.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 5

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at fifth grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade level vocabulary.
- Has an insufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) influence text.
- Inconsistently recognizes how story elements (e.g., plot, setting, characterization, theme) impact text.
- Inconsistently distinguishes stated main idea and relevant details in informational text.
- Inconsistently identifies and uses literary devices (e.g., simile, alliteration, metaphor, imagery).
- Inconsistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Inconsistently interprets informational text features (e.g., headings, maps, indexes).
- Inconsistently identifies defining characteristics of narrative and informational genres (e.g., poetry, myths, fantasies).
- Inconsistently answers literal, inferential, and critical questions with accuracy.

Grade 6

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor Grade 6
Advanced Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above sixth grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level. An *advanced* learner: • Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text. • Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary. • Has a thorough understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning and reliability of text. • Consistently identifies and analyzes how story elements (e.g., plot, setting, characterization, theme, point of view) impact text. • Consistently summarizes and analyzes informational text using stated and implied main idea and relevant details. • Consistently identifies and interprets literary devices (e.g., simile, alliteration, metaphor, imagery). • Consistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion). • Consistently interprets informational text features (e.g., headings, maps, indexes, charts). • Consistently distinguishes between defining characteristics of narrative and informational genres (e.g., poetry, myths, folk tales). • Consistently answers literal, inferential, critical, and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 6

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at sixth grade. A student scoring at the proficient level *generally* utilizes a variety of reading strategies to comprehend and interpret grade-level appropriate narrative and informational text.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a sufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning and reliability of text.
- Generally identifies and analyzes how story elements (e.g., plot, setting, characterization, theme, point of view) impact text.
- Generally summarizes and analyzes informational text using stated and implied main idea and relevant details.
- Generally identifies and interprets literary devices (e.g., simile, alliteration, metaphor, imagery).
- Generally applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Generally interprets informational text features (e.g., headings, maps, indexes, charts).
- Generally distinguishes between defining characteristics of narrative and informational genres (e.g., poetry, myths, folk tales).
- Generally answers literal, inferential, critical, and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 6

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at sixth grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has an insufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning of text.
- Inconsistently identifies how story elements (e.g., plot, setting, characterization, theme, point of view) impact text.
- Inconsistently distinguishes stated or implied main idea and relevant details in informational text.
- Inconsistently identifies and interprets literary devices (e.g., simile, alliteration, metaphor, imagery).
- Inconsistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion).
- Inconsistently interprets informational text features (e.g., headings, maps, indexes, charts).
- Inconsistently distinguishes between defining characteristics of narrative and informational genres (e.g., poetry, myths, folk tales).
- Inconsistently answers literal, inferential, critical, and interpretive questions with accuracy.

Grade 7

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor Grade 7
Advanced Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above seventh grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level. An *advanced* learner: • Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text. • Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary. • Has a thorough understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text. • Consistently identifies and analyzes how story elements (e.g., plot, setting, characterization, theme, point of view, conflict) impact text. • Consistently summarizes, analyzes, and synthesizes informational text using stated and implied main idea and relevant details. • Consistently analyzes author's use of literary devices (e.g., foreshadowing, personification, idiom, irony). • Consistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support). • Consistently interprets informational text features (e.g., headings, maps, indexes, charts, annotations). • Consistently makes inferences based on defining characteristics of narrative and informational genres (e.g., poetry, myths, folk tales, textbooks). • Consistently answers literal, inferential, critical and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 7

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at seventh grade. A student scoring at the proficient level *generally* utilizes a variety of reading strategies to comprehend and interpret grade-level appropriate narrative and informational text.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a sufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text.
- Generally identifies and analyzes how story elements (e.g., plot, setting, characterization, theme, point of view, conflict) impact text.
- Generally summarizes, analyzes, and synthesizes informational text using stated and implied main idea and relevant details.
- Generally analyzes author's use of literary devices (e.g., foreshadowing, personification, idiom, irony).
- Generally applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support).
- Generally interprets informational text features (e.g., headings, maps, indexes, charts, annotations).
- Generally makes inferences based on defining characteristics of narrative and informational genres (e.g., poetry, myths, folk tales, textbooks).
- Generally answers literal, inferential, critical, and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 7

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at seventh grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has an insufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning and reliability of text.
- Inconsistently identifies and analyzes how story elements (e.g., plot, setting, characterization, theme, point of view, conflict) impact text.
- Inconsistently summarizes informational text using stated main idea and relevant details.
- Inconsistently analyzes author's use of literary devices (e.g., foreshadowing, personification, idiom, irony).
- Inconsistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support).
- Inconsistently interprets informational text features (e.g., headings, maps, indexes, charts, annotations).
- Inconsistently makes inferences based on defining characteristics of narrative and informational genres (e.g., poetry, myths, folk tales, textbooks).
- Inconsistently answers literal, inferential, critical, and interpretive questions with accuracy and occasionally identifies supporting information in the text.

Grade 8

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor Grade 8
Advanced Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above eighth grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level. An *advanced* learner: • Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text. • Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary. • Has a thorough understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text. • Consistently identifies and analyzes how story elements (e.g., plot, setting, characterization, inferred and recurring theme, point of view, conflict) impact text. • Consistently summarizes, analyzes, and synthesizes informational text using stated and implied main idea and relevant details. • Consistently analyzes author's use of literary devices (e.g., foreshadowing, personification, idiom, irony, transitional devices). • Consistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support). • Consistently analyzes and evaluates information from text features (e.g., headings, maps, indexes, charts, annotations). • Consistently makes inferences based on defining characteristics of narrative and informational genres. • Consistently answers literal, inferential, critical and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 8

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at eighth grade. A student scoring at the proficient level *generally* utilizes a variety of reading strategies to comprehend and interpret grade-level appropriate narrative and informational text.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a sufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text.
- Generally identifies and analyzes how story elements (e.g., plot, setting, characterization, inferred and recurring theme, point of view, conflict) impact text.
- Generally summarizes, analyzes, and synthesizes informational text using stated and implied main idea and relevant details.
- Generally analyzes author's use of literary devices (e.g., foreshadowing, personification, idiom, irony, transitional devices).
- Generally applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support).
- Generally analyzes and evaluates information from text features (e.g., headings, maps, indexes, charts, annotations).
- Generally makes inferences based on defining characteristics of narrative and informational genres.
- Generally answers literal, inferential, critical, and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 8

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at eighth grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has an insufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text.
- Inconsistently identifies and analyzes how story elements (e.g., plot, setting, characterization, inferred and recurring theme, point of view, conflict) impact text.
- Inconsistently summarizes and analyzes informational text using stated main idea and relevant details.
- Inconsistently analyzes author's use of literary devices (e.g., foreshadowing, personification, idiom, irony, transitional devices).
- Inconsistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support).
- Inconsistently analyzes informational text features (e.g., headings, maps, indexes, charts, annotations).
- Inconsistently makes inferences based on defining characteristics of narrative and informational genres.
- Inconsistently answers literal, inferential, critical, and interpretive questions with accuracy and occasionally identifies supporting information in the text.

Grade 11

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor Grade 11
Advanced Overall student performance in reading reflects *high academic* performance on the standards and a *thorough* understanding of the content at or above eleventh grade. A student scoring at the advanced level *consistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level. An *advanced* learner: • Uses an on-grade-level or above-grade-level reading vocabulary to construct meaning from text. • Consistently applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary. • Has a thorough understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text. • Consistently analyzes and evaluates how story elements (e.g., plot, setting, characterization, inferred and recurring theme, point of view, conflict, mood) impact text. • Consistently summarizes, analyzes, synthesizes, and evaluates informational text using stated and implied main idea and relevant details. • Consistently analyzes author's use of stylistic and literary devices (e.g., foreshadowing, personification, irony, transitional devices, oxymoron, tone). • Consistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support, concept definition). • Consistently analyzes and evaluates information from text features (e.g., headings, maps, indexes, charts, annotations). • Consistently makes inferences based on defining characteristics of narrative and informational genres. • Consistently answers literal, inferential, critical, and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 11

Proficient

Overall student performance in reading reflects *satisfactory* performance on the standards and *sufficient* understanding of the content at eleventh grade. A student scoring at the proficient level *generally* utilizes a variety of reading strategies to comprehend and interpret grade-level appropriate narrative and informational text.

A *proficient* learner:

- Uses an on-grade-level reading vocabulary to construct meaning from text.
- Generally applies a variety of word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has a sufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text.
- Generally analyzes and evaluates how story elements (e.g., plot, setting, characterization, inferred and recurring theme, point of view, conflict, mood) impact text.
- Generally summarizes, analyzes, synthesizes, and evaluates informational text using stated and implied main idea and relevant details.
- Generally analyzes author's use of stylistic and literary devices (e.g., foreshadowing, personification, irony, transitional devices, oxymoron, tone).
- Generally applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support, concept definition).
- Generally analyzes and evaluates information from text features (e.g., headings, maps, indexes, charts, annotations).
- Generally makes inferences based on defining characteristics of narrative and informational genres.
- Generally answers literal, inferential, critical, and interpretive questions with accuracy and identifies supporting information in the text.

Nebraska State Accountability-Reading (NeSA-R) Performance Level Descriptor

Grade 11

Basic

Overall student performance in reading reflects *unsatisfactory* performance on the standards and *insufficient* understanding of the content at eleventh grade. A student scoring at the basic level *inconsistently* utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.

A *basic* learner:

- Uses a below-grade-level reading vocabulary to construct meaning from text.
- Inconsistently applies word-identification strategies (word structure, context, semantic relationships) to understand unfamiliar grade-level vocabulary.
- Has an insufficient understanding of how an author's purpose and perspective (beliefs, assumptions, biases) affect the meaning, reliability, and validity of text.
- Inconsistently analyzes and evaluates how story elements (e.g., plot, setting, characterization, inferred and recurring theme, point of view, conflict, mood) impact text.
- Inconsistently summarizes, analyzes, and synthesizes informational text using stated and implied main idea and relevant details.
- Inconsistently analyzes author's use of literary devices (e.g., foreshadowing, personification, irony, transitional devices, oxymoron, tone).
- Inconsistently applies knowledge of organizational patterns of informational text (e.g., sequence cause/effect, fact/opinion, proposition/support, concept definition).
- Inconsistently analyzes and evaluates information from text features (e.g., headings, maps, indexes, charts, annotations).
- Inconsistently makes inferences based on defining characteristics of narrative and informational genres.
- Inconsistently answers literal, inferential, critical, and interpretive questions with accuracy and occasionally identifies supporting information in the text.

Appendix B: Meeting Agenda

Appendix B.1 Agenda

NeSA-R

Nebraska Bookmark Standard Setting Meeting

Sunday June 27, 2010

Hotel Check-in for those traveling long distances

Monday June 28, 2010 *(times are approximate depending on work completion)*

8:00 – 8:30 Breakfast and Check-in

8:30 – 10:30 Training in Large Group in Room E&F

10:35 – 12:00 Grade Group Breakouts

Reading Grade	*Teachers who teach*	*Room*
4	*Grades 3, 4, 5*	*B*
7	*Grades 6, 7, 8*	*C*
11	*Grades 10, 11, 12*	*D*

12:00 – 1:00 Lunch in Lancaster 4, 5, 6

1:00 – Completion Complete work for first Grade Group

Tuesday June 29, 2010 *(times are approximate depending on work completion)*

8:00 – 8:30 Breakfast and Check-in

8:30 – 9:00 Review Monday in Large Group Room E&F

9:00– 12:00 Meeting in Small Groups by Grade

Reading Grade	*Teachers who teach*	*Room*
3	*3, 4, 5*	*B*
8	*6, 7, 8 and 10 +*	*C,D*

12:00 – 1:00 Lunch in Lancaster 4, 5, 6

1:00 – Completion Continue in Small Groups by Grade

Wednesday June 30, 2010 *(times are approximate depending on work completion)*

8:00 – 8:30 Breakfast and Check-in

8:30 – 12:00 Meeting in Small Group for grades 5 and 6

Reading Grade	*Teachers who teach*	*Room*
5	*3, 4, 5*	*TBD*
6	*6, 7, 8*	*TBD*

12:00 – 1:00 Lunch in Lancaster

1:00 – Completion Continue in Small Groups

Appendix B.2: Groupings and Room Assignments

Reading	Monday			Tuesday		Wednesday	
June 28-30, 2010	**Room 1 (room for 45)**	**Room 2 (room for 45)**	**Room 3 (room for 30)**	**Room 1 (room for 45)**	**Room 2 (room for 60)**	**Room 1 (room for 45)**	**Room 2 (room for 45)**
8:00 AM	Breakfast			Breakfast		Breakfast	
8:15 AM							
8:30 AM	Training Large Group			Presentation of Results from previous day		**Grade 5** Take Test	**Grade 6** Take Test
8:45 AM							
9:00 AM						PLD Review	PLD Review
9:15 AM							
9:30 AM				**Grade 3** Take test	**Grade 8** Take test		
9:45 AM						R1 OIB review and Bookmark Placement	
10:00 AM							
10:15 AM				PLD review	PLD review		
10:30 AM	Move to grade level rooms					Break and Analysis	
10:45 AM	**Grade 4** Take test	**Grade 7** Take test	**Grade 11** Take test			R1 Feedback and Discussion	
11:00 AM				R1 OIB review and Bookmark Placement			
11:15 AM	PLD review	PLD review	PLD review			R2 bookmark Adjustments	
11:30 AM							
11:45 AM							
12:00 PM	Lunch and Analysis			Lunch and Analysis		Lunch and Analysis	
12:15 PM							
12:30 PM							
12:45 PM							
1:00 PM	R1 OIB review and Bookmark placement			R1 Feedback and Discussion		R2 Feedback and Discussion	
1:15 PM							
1:30 PM						R3 Bookmark Adjustments	
1:45 PM							
2:00 PM	Break and Analysis			R2 Bookmark Adjustments			
2:15 PM	R1 Feedback and Discussion						
2:30 PM							
2:45 PM							
3:00 PM	R2 Bookmark Adjustments			Break and Analysis			
3:15 PM				R2 Feedback and Discussion Adding in NAEP data as available for Grade 8			
3:30 PM							
3:45 PM	Break and Analysis						
4:00 PM	R2 Feedback and Discussion Adding in NAEP and ACT data as available						
4:15 PM				R3 Bookmark Adjustments			
4:30 PM							
4:45 PM							
5:00 PM		R3					

Appendix C: PowerPoint: Setting Academic Proficiency Standards

Setting Academic Proficiency Standards for the Nebraska State Accountability Reading Assessment(NeSA-R)
June 28-30, 2010

Welcome and Introductions

- Dr. Pat Roschewski
  - Director of Statewide Assessment, Nebraska Department of Education
- John Moon, Department of Education
- Jan Hoegh, Department of Education

2

Introduction of DRC Staff

- David Chayer, Trainer and Lead Facilitator
- Ronald Mead, Group Facilitator
- Vince Primoli, Group Facilitator
- Julie Korts, Data Analyst
- Jessica Tickle, Content Specialist

3

Introduction of DRC Staff (cont.)

- Project Management
  - John Born
  - Valerie Cook

4

Logistics: Date, Panelists, Method

- Date
  - June 28-30, 2010
- Panelists
  - Approximately 15 per grade
  - Selected to span grades 3 through 8, 11 and Higher Education
  - Grade Groups may be adjusted depending on shortages in certain grade groups
- Method
  - Modified Bookmark

5

Forms and Documentation

- Panelist Experience Survey forms
- Reimbursement forms
- Confidentiality forms
- Task Readiness forms
- Bookmark Meeting Evaluation forms

6

Purpose of the Meeting

- To set NeSA-R cutscores that categorize students into one of three achievement levels:
  - Advanced
  - Proficient
  - Basic
- To articulate these expectations across grades 3 through 8 and 11

7

Schedule

- Monday:
  - Grade 4 (panelists from grades 3, 4, 5)
  - Grade 7 (panelists from grades 6, 7, 8)
  - Grade 11 (panelists from grades 10, 11, 12 and Higher Education)

8

Schedule

- Tuesday:
  - Grade 3 (panelists from grades 3, 4)
  - Grade 6 (panelists from grades 5, 6, 7)
  - Grade 8 (panelists from grades 8, 10+)

9

Schedule

- Wednesday:
  - Grade 5 (panelists from grades 4, 5, 6)
  - Articulation of results (grades 3 to 8, 11)

10

Methodology

Modified Bookmark

- One in a broad category of methods commonly referred to as item mapping that focuses on items rather than examinees
- Places emphasis on what a student should know and be able to do

11

Step 1: How do students demonstrate their proficiency?

- Take the test
  - Provides panelists a feel for students' testing experience on the operational administration of the NeSA Reading

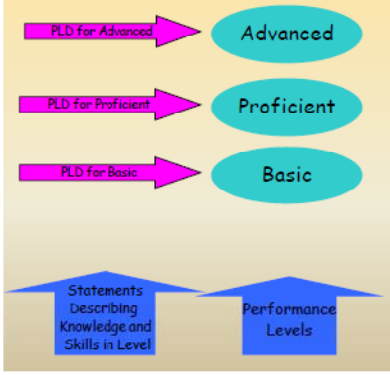
12

Step 2: What do the performance levels mean?

- Performance Level Descriptors (PLDs) are:
  - Statements that describe the knowledge and skills expected at each of the three achievement levels
  - Unique to each grade and subject
  - **Middle** of the level; not the borderline students

13

The Relationships Between PLDS and Performance Levels

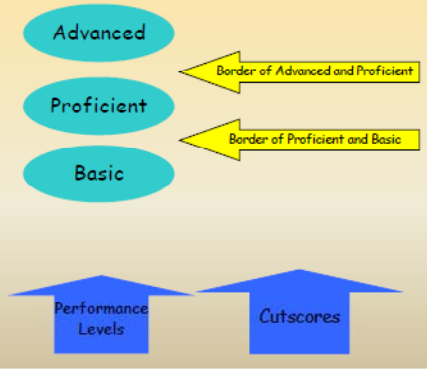


Step 3: What Defines Borderline Students?

- Visualize Nebraska students who are:
  - Just barely leaving one level, and
  - Just barely entering the next higher level

15

The Relationships Between Performance Levels and Cutscores

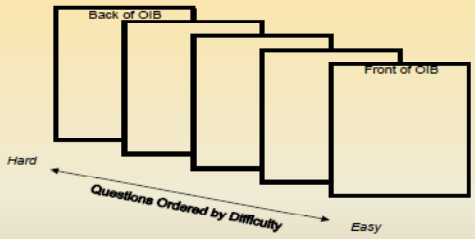


Step 4: What are we expected to do?

- Panelists are presented with operational test questions ordered from easiest to most difficult

17

Ordered Item Booklet (OIB)



•Based on Spring 2010 data

18

Step 5: Where do we draw the line?

- Panelists are asked to make judgments about which items students at the borderline between two performance levels are able to get correct and which ones they are not.
  - Criterion: "67 or more out of 100"
- Place the "bookmark" on the first item that does not meet the criterion.

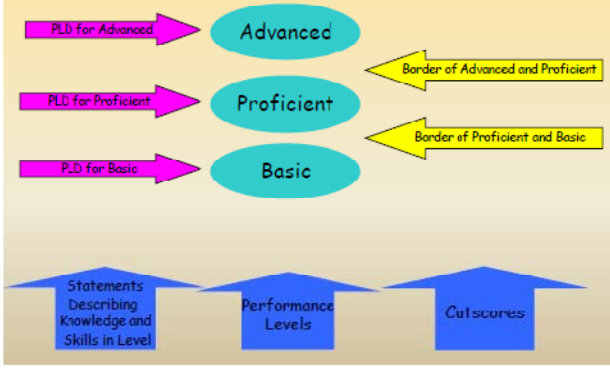
19

What happens to our bookmarks?

- These collective judgments determine the recommended cutscores that separate:
  - ***Proficient*** students from ***Basic*** students
  - ***Advanced*** students from ***Proficient*** students

20

The Relationships Among PLDs, Performance Levels, and Cutscores



Overview of Process

- Three rounds of individual judgments
  - Group discussions
  - Opportunities to revise judgments
  - Data will be presented at the beginning of Rounds 2 and 3
    - For example, the percent of students that would fall into each of the three performance levels based on the group recommendation from the previous round

22

Articulation Across Grades The Process

- Results from Rounds 1, 2, and 3 will be presented for all grades as they become available
- NAEP results for Nebraska students in 2009 from grades 4 and 8 will be presented in the associated grades
  - After Round 2
- ACT College Readiness scores will be presented for grade 11
  - After Round 2

23

Panelists' Roles and Responsibilities

- Satisfy yourself that you have contributed to a group recommendation that is based on your experience and professional judgment

24

Outcomes

- The recommendations from this meeting will be presented to the State Board for review, along with other relevant information
- Final cutscores will be established and approved by the State Board

25

Test Security

- Check in
  - Distribution of training materials
- Security
  - ALL materials must remain in the room.
  - For unscheduled breaks, please notify the Room Facilitator.
  - No discussion related to any of the secure materials outside of the rooms, including breaks, lunch, and dinner
- Check out
  - Turn in materials at the end of each day.
  - Materials will be returned to you at the start of each day

26

Courtesy Reminders

- Cell phones:
  - Please turn off or set to silent
  - If you must take a call, please excuse yourself from the room quietly (leave all secure materials in the room)
- E-mail, PDAs, Blackberry, or other computer work:
  - Please refrain except during extended breaks
- Conversations:
  - Please be considerate of others

27

Return Times

- Lead Facilitator will indicate when you should return (e.g., after breaks and lunch)
- Times may be different for each panel

28

Training Materials

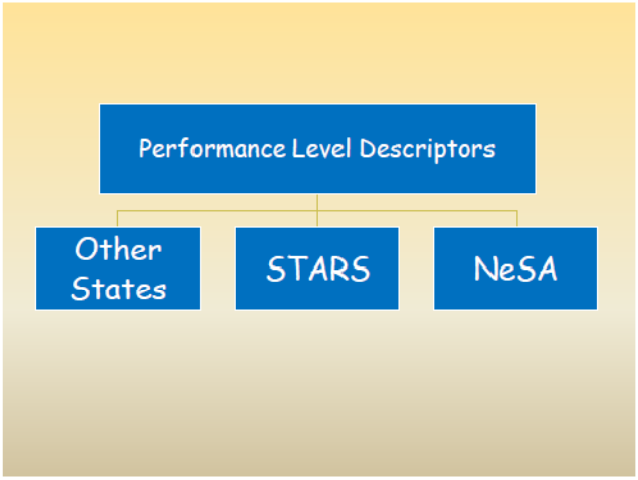
- Sample:
  - NeSA-R Performance Level Descriptors (grade 4 reading)
  - Five items (NAEP grade 4 reading)
  - Item Map
  - Item Separation Chart
  - Participant Rating Form
  - Results

29

Performance Level Descriptors

•Performance level descriptors (PLD's) describe the level of knowledge and skills required at each performance level.

•Basic
•Proficient
•Advanced



NeSA-R Performance Level Descriptors

- Two levels
- Policy definitions provide descriptors for each level of proficiency
- Policy definitions with full descriptors communicate content expectations for vocabulary and comprehension at each grade level assessed

Performance Level Descriptors
Policy Statements

Advanced	Overall student performance in reading reflects high academic performance on the standards and a thorough understanding of the content at or above third grade. A student scoring at the advanced level consistently utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at or above grade level.
Proficient	Overall student performance in reading reflects satisfactory performance on the standards and sufficient understanding of the content at third grade. A student scoring at the proficient level generally utilizes a variety of reading skills and strategies to comprehend and interpret narrative and informational text at grade level.
Basic	Overall student performance in reading reflects unsatisfactory performance on the standards and insufficient understanding of the content at third grade. A student scoring at the basic level inconsistently utilizes a variety of reading skills and strategies to comprehend and interpret narrative and information text at grade level.

Performance Level Descriptors
Full Descriptors

Advanced	• Consistently answers literal and inferential questions with accuracy and provides supporting information.
Proficient	• Generally answers literal and inferential questions with accuracy.
Basic	• Inconsistently answers literal questions with accuracy.

Bookmark Training
(cont.)

- The Item Map

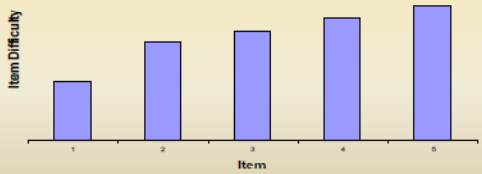
Sample Item Map

OIB Page	Item Type	Key	Standard	Round 1	Round 2	Round 3
1	MC	D	3			
2	MC	C	2			
3	MC	D	5			
4	MC	A	4			
5	MC	B	1			

35

Bookmark Training
(cont.)

- Item Separation Chart



36

Placing the Bookmark: The Region of Uncertainty

- Identify ***groups*** of items that are probable choices based on your judgment
- Do not focus on a single item
- Utilize the "item separation chart"

37

Bookmark Training (cont.)

- Participant Rating Form
  - For the formal part of the meeting, two bookmarks will be placed in the OIB
    - 1st: Basic/Proficient
    - 2nd: Proficient/Advanced

38

NEBRASKA STATE ACCOUNTABILITY – READING (NESA-R) STANDARD SETTING

TRAINING PARTICIPANT RATING FORM

Name: _____

Grade: _____

Number: _____

OIB Page Number of Bookmark		
ROUND	BASIC/PROFICIENT	PROFICIENT/ADVANCED
Round 1		
Round 2		
Round 3		

39

Training (cont.)

- Placing a Bookmark
  - Go through the OIB page by page and assess whether a borderline ***Proficient*** student, according to the PLDs, has a sufficient probability of answering each item correctly
    - For multiple-choice (MC) items, ***sufficient*** is .67

40

Training (cont.)

- Does a borderline **Proficient** student have at least a .67 probability of answering this item correctly?
- If yes, turn the page and make the same judgment about the next most difficult item

41

Training (cont.)

- Continue until you reach an item that the borderline ***Proficient*** student would **not** have a .67 chance of answering correctly
- ***Key Point***
  - Place your post-it bookmark on this page, which represents the first item that the borderline student would **not** answer correctly 67 percent of the time
    - ***Reminder:*** Region of Uncertainty!
  - Record this page number on your Rating Form

42

Practice Exercise

43

Practice Steps

- Review the five sample reading questions
- Review the NeSA-R PLDs for Proficient and Basic
- Visualize a student just barely out of the Basic level and just barely into the Proficient level
- Indicate on the Sample Rating Form the first item you judge your borderline student would get correct ***less*** than 67 percent of the time

44

Done?

45

Show of Hands!

46

Sample Results

Round 1 Bookmark Placements by Panelist and Total

Name	Basic/Proficient	Proficient/Advanced
Participant1	6	14
Participant2	11	17
Participant3	5	17
Participant4	8	17
Participant5	12	27
Participant6	2	10
Participant7	6	27
Participant8	10	25
Participant9	33	34
Participant10	8	32
Participant11	8	28
Table Median	8	25

47

Sample Results

Round 1 Ratings and Impacts

Raw Score	Ability	SE	Basic/Proficient	Proficient/Advanced	Cum. %	
.	.	.			.	
.	.	.			.	
.	.	.			.	
20	646	24			4.47	
21	655	23			5.93	
22	661	22			6.3	
23	667	21			7.35	
24	673	20			8.42	
25	679	20			9.9	
26	685	20			11.29	
27	691	19	P(1,4)		13.01	
28	697	19	P(1,10)		14.9	
29	702	19	P(1,5)P(1,6)P(1,7)P(1,13)P(1,14)		16.84	Basic
30	708	19	P(1,1)P(1,3)P(1,9)P(1,16)P(1,15)		18.19	Proficient
31	714	19			21.64	
32	720	19			24.43	
33	726	19	P(1,2)		27.46	
34	732	19	P(1,8)P(1,11)		30.82	
35	738	19	P(1,12)	P(1,16)	34.57	
36	745	19	P(1,13)	P(1,7)	38.48	
37	751	19			42.65	
38	758	19		P(1,3)P(1,10)P(1,14)P(1,15)	47.19	Proficient
39	765	20		P(1,1)P(1,2)P(1,3)P(1,4)P(1,5)P(1,6)	51.99	Advanced
40	772	20			57.17	
41	780	21			62.77	
42	788	22		P(1,8)P(1,13)	68.46	
43	797	24		P(1,12)P(1,11)	74.13	
44	807	27			79.5	
45	819	31			85.59	
46	833	36			90.79	
47	852	43			94.09	
48	878	46			96.05	
49	926	47			98.58	
50	990	47			100	

48

Sample Results (cont.)

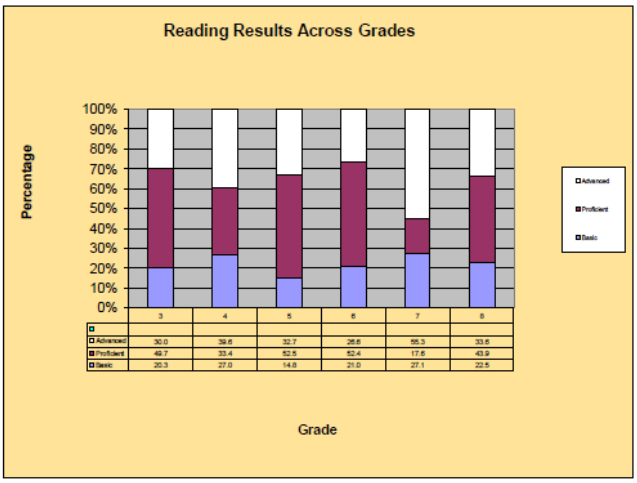
Percent in Performance Level

Round 1		
Basic	Proficient	Advanced
33.4	51.2	15.4

Round 2		
Basic	Proficient	Advanced
30.5	47.2	22.3

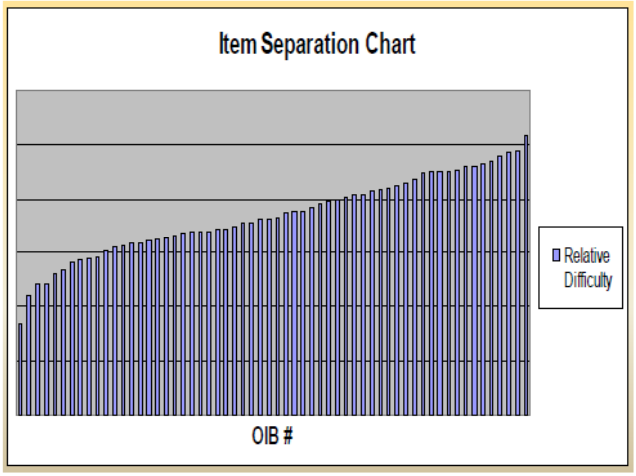
Round 3		
Basic	Proficient	Advanced
36.4	41.7	21.9

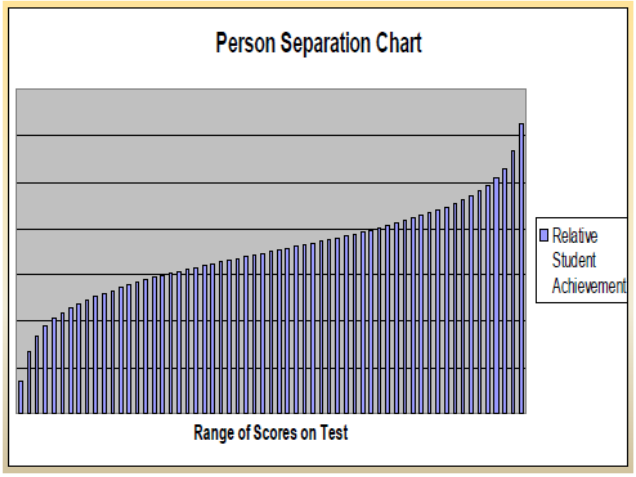
49

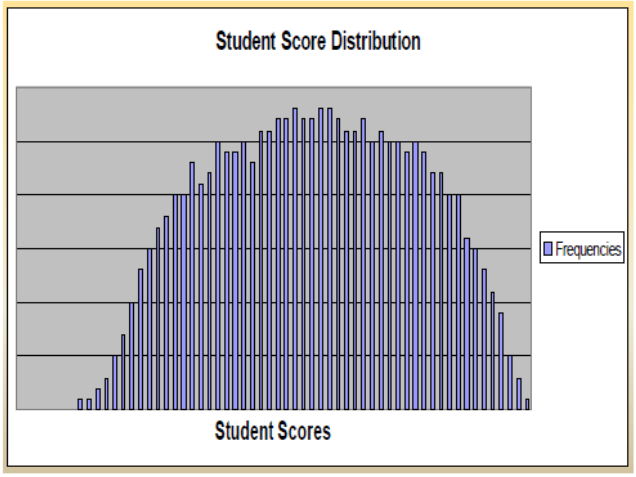


Psychometrics: An Introduction

51







What You May And May Not Discuss Outside Of This Meeting

- You may discuss:
  - The processes used
- You may **not** discuss:
  - The results
  - The contents of the secure materials
    - Items
    - PLDs

55

Questions?

56

What's Next?

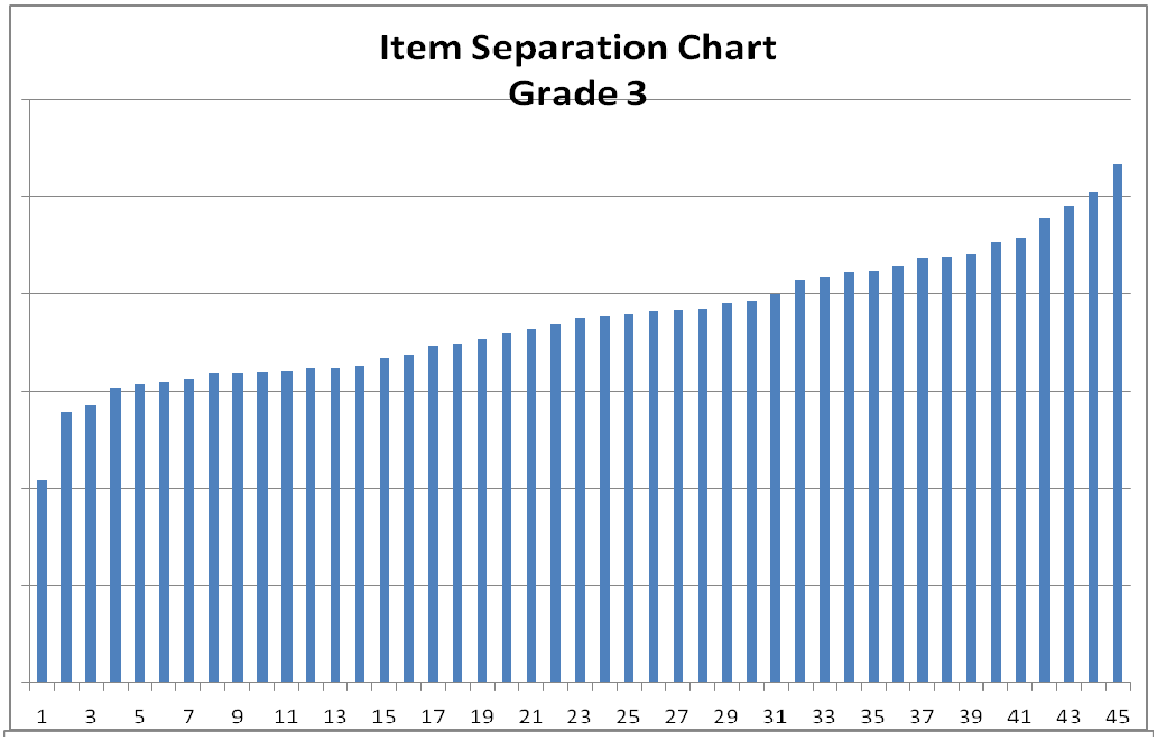
- Tuesday and Wednesday morning
  - Breakfast starts at 8:00 am
  - Check-in from 8:00-8:25 am
  - Meeting begins at 8:30 am

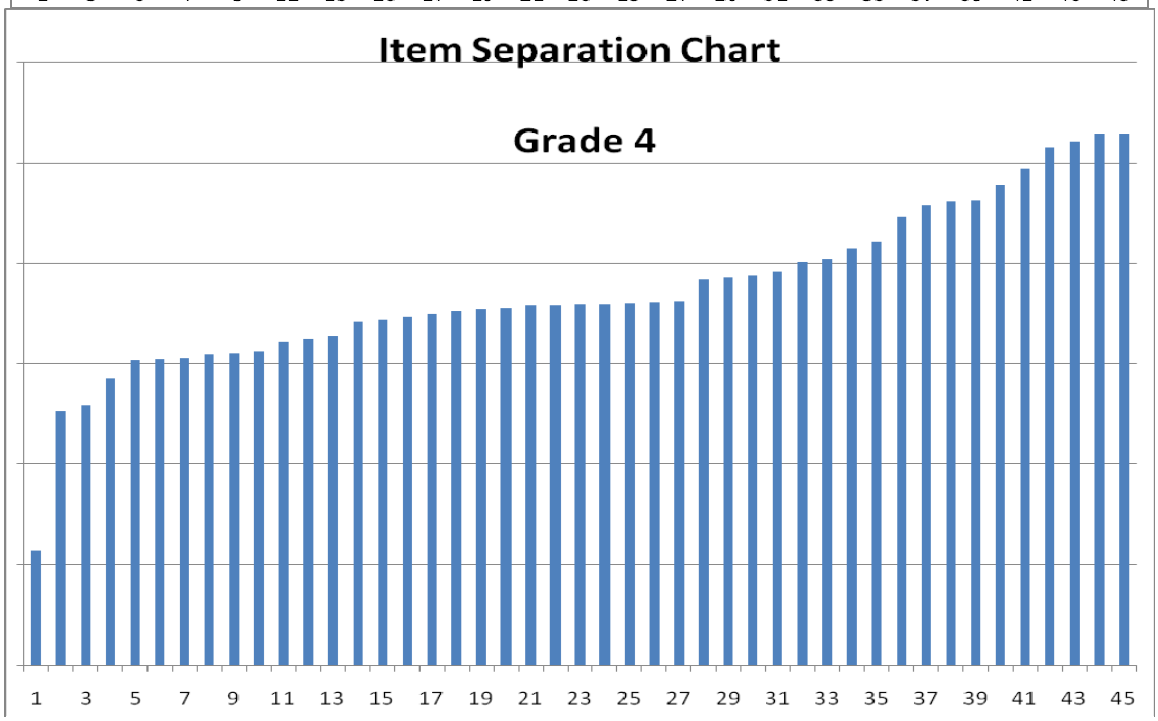
57

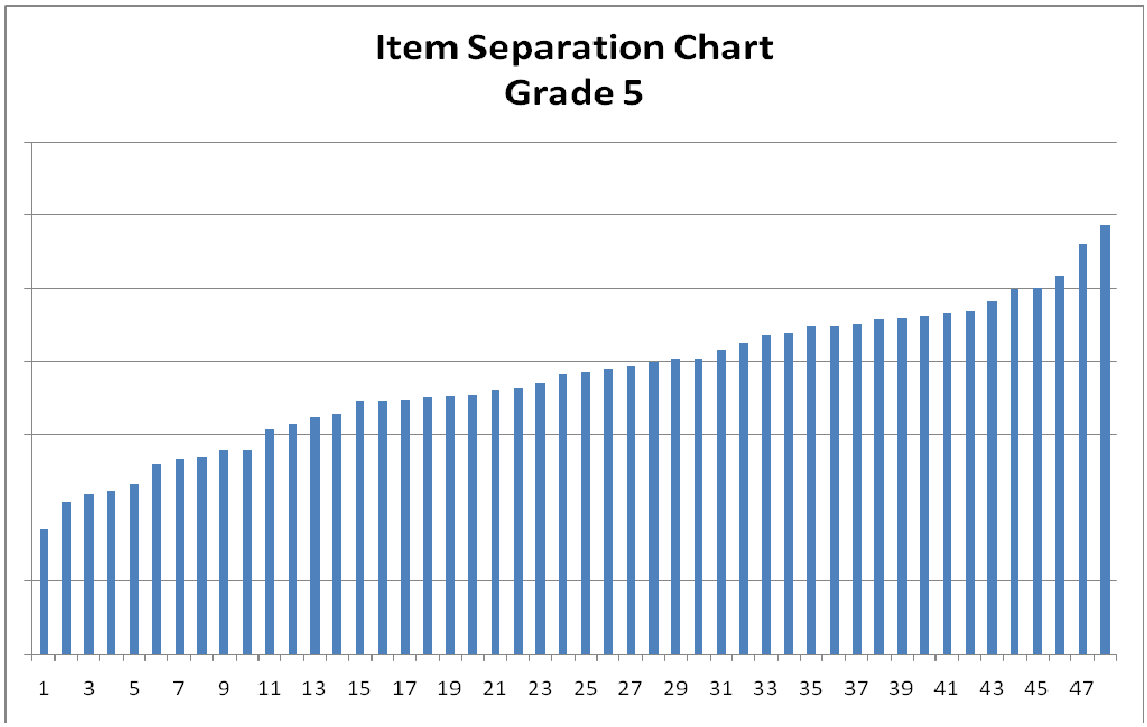
Appendix D: Impacts by Round

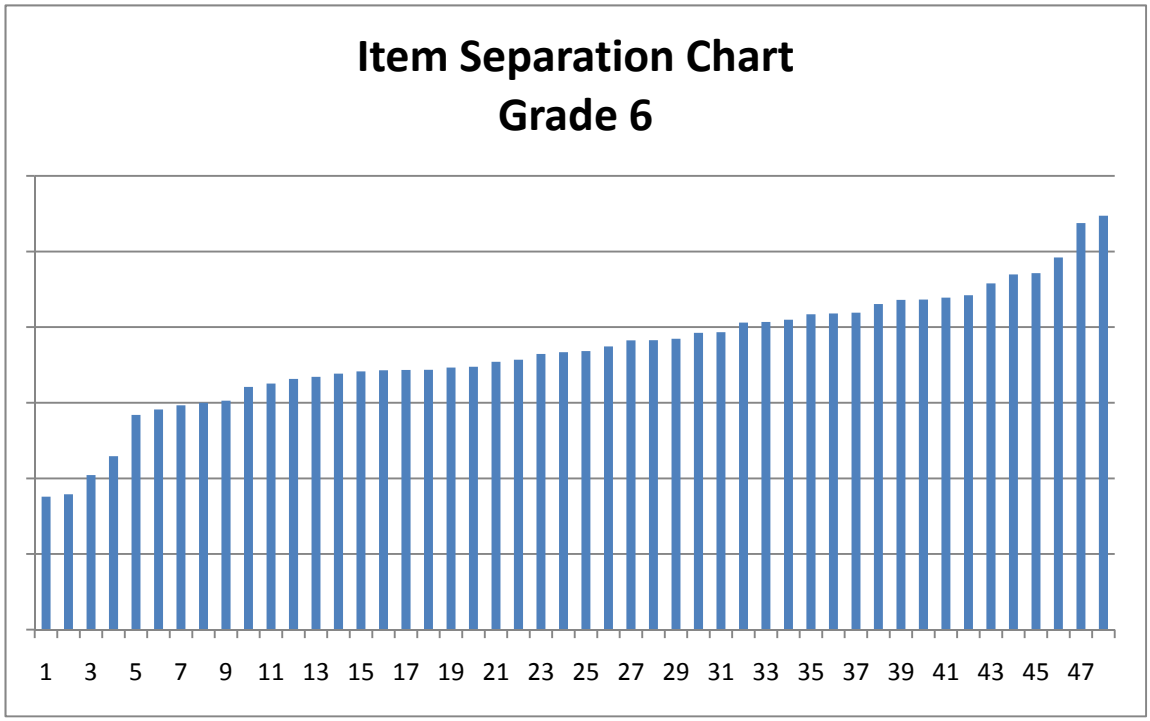
Reading	Below the Standards	Meets the Standards	Exceeds the Standards
Grade 3			
Round 1	19.4	27.2	53.4
Round 2	19.4	31.1	49.5
Round 3	19.4	35.5	45.1
Grade 4			
Round 1	16.7	25.2	58.1
Round 2	14.8	27.1	58.1
Round 3	14.8	46.7	38.5
Grade 5			
Round 1	15.4	44.0	40.6
Round 2	15.4	44.0	40.6
Round 3	15.4	44.0	40.6
Grade 6			
Round 1	18.6	33.6	47.8
Round 2	20.7	36.6	42.7
Round 3	20.7	36.6	42.7
Grade 7			
Round 1	22.3	36.5	41.2
Round 2	15.2	43.6	41.2
Round 3	22.3	36.5	41.2
Grade 8			
Round 1	24.0	33.5	42.5
Round 2	24.0	38.4	37.6
Round 3	24.0	38.4	37.6
Grade 11			
Round 1	22.8	28.3	48.9
Round 2	22.8	33.1	44.1
Round 3	22.8	43.5	33.7

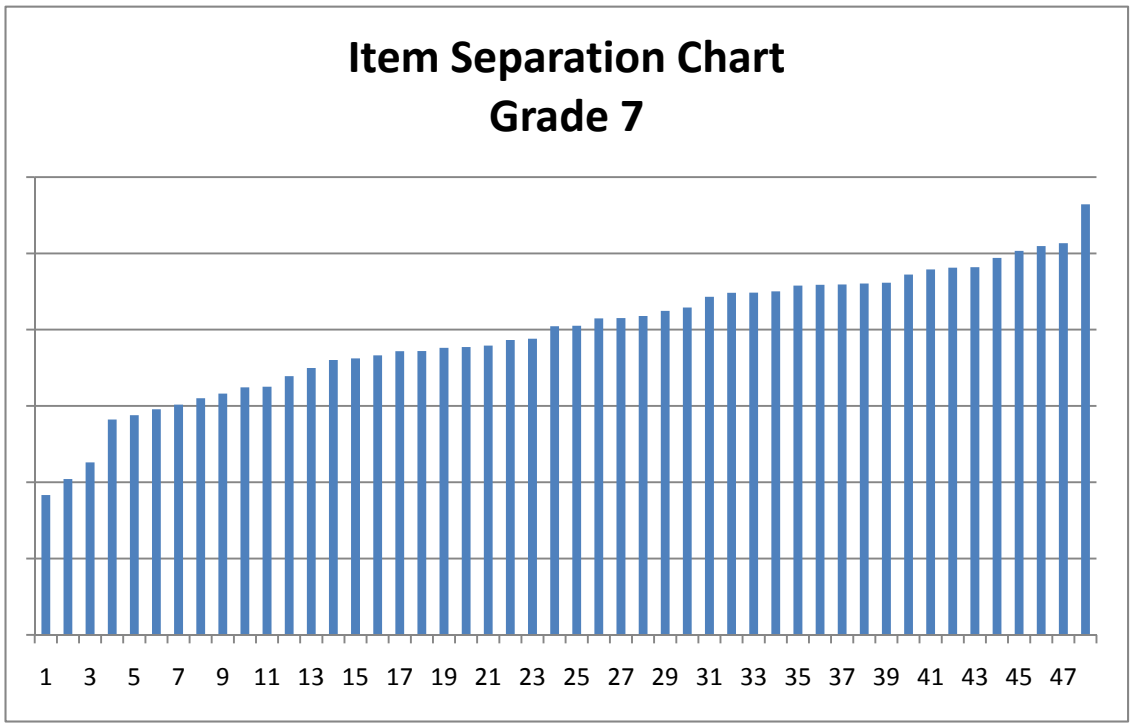
Appendix E: Item Separation Maps

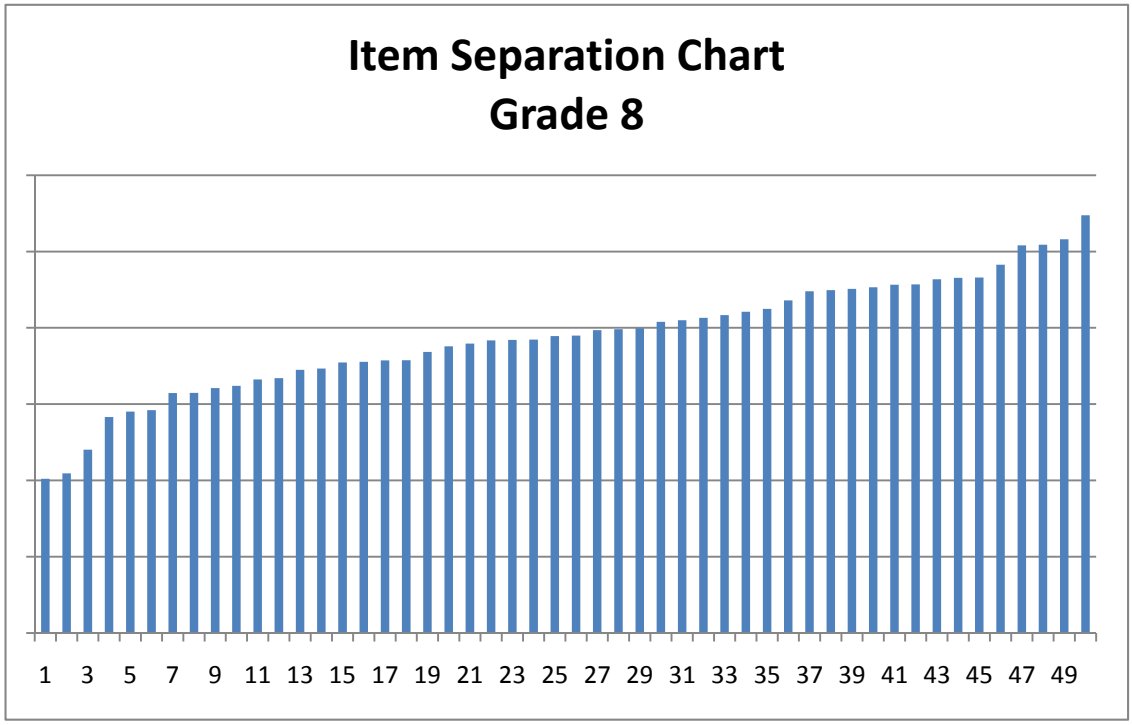


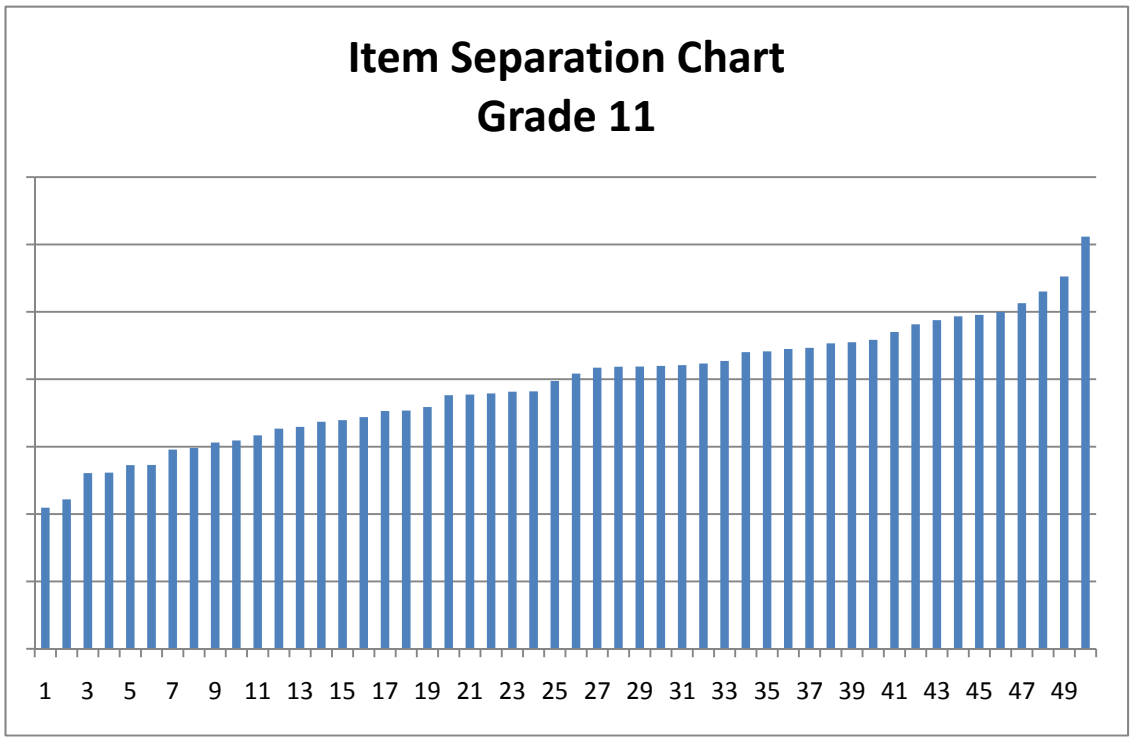



Item Separation Chart
Grade 5
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47


Item Separation Chart
Grade 6
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47


Item Separation Chart
Grade 7
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47


Item Separation Chart
Grade 8
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


Item Separation Chart
Grade 11
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49

Appendix F: Contrasting Groups Summaries

Table F.1: Overall Contrasting Group Summary Data

Group	Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 11	
	State	Teacher Rated	State	Teacher Rated	State	Teacher Rated	State	Teacher Rated	State	Teacher Rated	State	Teacher Rated	State	Teacher Rated
Student Count														
Total	21553	1424	21185	1437	20751	1096	20483	1200	20387	991	20400	1262	20542	1407
Gender														
Male	11010	716	10859	754	10612	553	10515	631	10451	502	10397	632	10403	705
Female	10543	708	10326	683	10139	543	9968	569	9936	489	10003	630	10139	702
Ethnicity														
African Amer.	1787	67	1745	63	1643	31	1595	31	1594	41	1572	32	1324	32
Amer. Indian	448	17	424	15	385	16	323	11	359	7	334	8	292	13
Hispanic	3335	204	3194	216	3071	188	2929	178	2886	146	2752	165	2276	114
Asian	481	36	469	21	492	17	432	18	437	9	442	22	424	29
White	15502	1100	15353	1122	15160	844	15204	962	15111	788	15300	1035	16226	1219
Teacher Rating														
Basic		521		497		366		338		306		324		367
Proficient		644		669		486		561		434		576		669
Advanced		259		271		244		301		251		362		371
Performance Level--Final														
Basic	6998	416	6458	413	6766	341	6510	347	6308	287	6029	354	6453	377
Proficient	10231	701	10181	748	9993	553	9945	588	9792	459	9829	633	10348	727
Advanced	4324	307	4546	276	3992	202	4028	265	4287	245	4542	275	3741	303
Correlation														
		0.613		0.595		0.626		0.626		0.642		0.651		0.593

Table F.2: Agreement between Teacher Ratings and Final Performance Level Status

Gr 3		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**316**	98	2
	Proficient	194	**410**	97
	Advanced	11	136	**160**

Gr 4		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**306**	99	8
	Proficient	183	**452**	113
	Advanced	8	118	**150**

Gr 5		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**249**	86	6
	Proficient	112	**331**	110
	Advanced	5	69	**128**

Gr 6		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**238**	106	3
	Proficient	92	**363**	133
	Advanced	8	92	**165**

Gr 7		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**201**	83	3
	Proficient	102	**264**	93
	Advanced	3	87	**155**

Gr 8		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**234**	109	11
	Proficient	87	**398**	148
	Advanced	3	69	**203**

Gr 11		Teacher Rating		
		Basic	Proficient	Advanced
Actual Performance	Basic	**250**	120	7
	Proficient	109	**438**	180
	Advanced	8	111	**184**

Table F.3: Subgroup Summary by Grade

Grade 3

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		21553	32.6	8.6	0.91	100.9	36.4	32.5	47.5	20.1
Gender	Male	11010	32.0	8.8	0.91	98.5	36.6	35.5	45.5	19.0
	Female	10543	33.2	8.2	0.90	103.4	35.9	29.3	49.5	21.2
Ethnicity	African American	1787	28.3	8.9	0.90	83.9	33.3	52.6	37.7	9.7
	American Indian	448	26.0	9.2	0.90	75.8	33.1	62.1	31.9	6.0
	Hispanic	3335	28.5	8.5	0.89	83.8	31.3	51.5	41.1	7.3
	Asian	481	33.1	9.3	0.93	104.9	40.9	30.1	46.2	23.7
	White	15502	34.1	7.9	0.90	107.2	35.5	25.3	50.4	24.3
Special Ed	No	18208	33.4	8.1	0.90	104.3	35.6	28.6	49.3	22.1
	Yes	3345	27.8	9.5	0.91	82.7	35.1	53.6	37.4	9.0
ELL	No	19671	33.2	8.4	0.91	103.3	36.3	29.6	48.7	21.6
	Yes	1882	26.6	8.1	0.87	76.6	27.6	62.2	34.2	3.6
FLS	No	10915	35.1	7.5	0.90	111.6	35.4	20.9	51.3	27.9
	Yes	10638	30.0	8.8	0.90	90.0	34.1	44.4	43.6	12.1

Grade 4

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		21185	32.6	7.9	0.89	103.8	41.1	30.5	48.1	21.5
Gender	Male	10859	32.0	8.1	0.89	100.7	41.4	33.3	47.1	19.5
	Female	10326	33.3	7.6	0.89	107.0	40.5	27.5	49.0	23.5
Ethnicity	African American	1745	28.2	8.6	0.89	81.6	39.5	51.9	39.0	9.1
	American Indian	424	27.5	8.1	0.88	78.0	36.3	57.3	37.7	5.0
	Hispanic	3194	29.1	7.9	0.87	85.2	36.9	48.6	42.3	9.1
	Asian	469	34.0	8.1	0.91	112.4	44.0	22.6	48.4	29.0
	White	15353	34.0	7.3	0.88	110.6	39.7	23.8	50.6	25.7
Special Ed	No	17663	33.7	7.3	0.88	108.7	39.3	25.4	50.7	23.9
	Yes	3522	27.5	8.9	0.90	79.1	40.8	56.1	34.8	9.1
ELL	No	19515	33.1	7.7	0.89	106.3	40.8	27.9	49.2	23.0
	Yes	1670	26.8	7.5	0.85	74.5	32.7	61.1	35.0	3.8
FLS	No	10856	34.9	6.9	0.87	115.3	38.8	19.6	51.3	29.1
	Yes	10329	30.3	8.2	0.89	91.6	39.9	42.0	44.6	13.4

Grade 5

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		20751	34.2	8.0	0.88	101.0	41.5	32.6	48.2	19.2
Gender	Male	10612	33.8	8.2	0.89	99.0	42.0	34.7	47.1	18.3
	Female	10139	34.7	7.8	0.88	103.1	40.9	30.5	49.3	20.2
Ethnicity	African American	1643	29.7	8.7	0.89	78.7	40.9	55.1	35.8	9.1
	American Indian	385	28.7	8.8	0.89	74.1	40.8	59.2	34.5	6.2
	Hispanic	3071	30.2	8.0	0.87	80.2	37.4	52.7	40.7	6.6
	Asian	492	36.1	8.4	0.91	113.3	46.0	25.6	43.5	30.9
	White	15160	35.6	7.3	0.87	107.9	39.7	25.7	51.5	22.8
Special Ed	No	17514	35.3	7.3	0.87	106.4	39.4	27.2	51.2	21.6
	Yes	3237	28.2	8.8	0.89	71.9	40.5	61.7	31.9	6.4
ELL	No	19423	34.7	7.8	0.88	103.6	40.8	29.9	49.7	20.4
	Yes	1328	26.6	7.6	0.84	63.6	32.8	72.2	26.1	1.7
FLS	No	10748	36.5	7.0	0.86	112.9	39.1	21.4	52.1	26.5
	Yes	10003	31.7	8.2	0.88	88.2	40.2	44.7	43.9	11.4

Grade 6

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		20483	35.1	8.3	0.90	101.3	41.2	31.8	48.6	19.7
Gender	Male	10515	34.4	8.6	0.90	97.6	41.6	35.2	47.3	17.6
	Female	9968	35.9	7.9	0.89	105.3	40.4	28.2	49.9	21.9
Ethnicity	African American	1595	31.0	9.0	0.89	81.4	39.4	50.6	41.3	8.2
	American Indian	323	29.0	9.7	0.91	73.1	41.9	58.8	34.1	7.1
	Hispanic	2929	31.1	8.7	0.89	81.4	38.5	51.8	40.4	7.9
	Asian	432	36.1	8.7	0.91	107.2	43.7	26.9	48.6	24.5
	White	15204	36.4	7.7	0.88	107.7	39.6	25.5	51.2	23.3
Special Ed	No	17411	36.5	7.4	0.87	107.5	38.4	25.5	52.2	22.3
	Yes	3072	27.5	9.2	0.89	66.4	38.6	67.6	27.8	4.5
ELL	No	19422	35.6	8.1	0.89	103.5	40.5	29.5	49.8	20.6
	Yes	1061	26.8	8.3	0.86	62.6	33.6	72.7	25.4	2.0
FLS	No	10804	37.4	7.3	0.88	112.5	38.8	21.1	52.4	26.5
	Yes	9679	32.6	8.7	0.89	88.9	40.2	43.7	44.2	12.1

Grade 7

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		20387	33.1	8.4	0.88	104.3	39.4	30.9	48.0	21.0
Gender	Male	10451	32.4	8.6	0.89	101.0	40.0	34.0	47.0	19.0
	Female	9936	33.8	8.0	0.88	107.7	38.5	27.7	49.1	23.1
Ethnicity	African American	1594	27.0	8.7	0.88	77.0	37.1	60.2	32.9	6.8
	American Indian	359	27.9	8.4	0.87	80.7	36.2	55.4	38.2	6.4
	Hispanic	2886	28.9	8.5	0.87	85.1	37.0	50.3	41.4	8.3
	Asian	437	33.7	9.4	0.91	108.8	45.1	27.7	43.7	28.6
	White	15111	34.6	7.6	0.87	111.2	37.2	23.7	51.3	25.1
Special Ed	No	17598	34.3	7.6	0.87	109.7	37.2	25.1	51.2	23.8
	Yes	2789	25.3	8.4	0.86	69.8	35.0	67.9	28.3	3.8
ELL	No	19550	33.5	8.2	0.88	105.9	38.8	29.2	49.0	21.8
	Yes	837	24.3	8.0	0.84	65.4	32.8	72.6	25.2	2.2
FLS	No	10963	35.7	7.2	0.86	116.3	36.6	18.8	51.7	29.4
	Yes	9424	30.1	8.6	0.88	90.2	37.9	45.0	43.7	11.3

Grade 8

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		20400	35.2	8.6	0.89	102.5	37.8	29.6	48.2	22.3
Gender	Male	10397	34.1	8.9	0.89	97.4	37.9	34.2	47.5	18.3
	Female	10003	36.5	8.1	0.88	107.9	36.9	24.7	48.9	26.4
Ethnicity	African American	1572	29.8	9.1	0.89	79.2	35.4	54.1	38.8	7.1
	American Indian	334	30.8	8.8	0.88	83.3	35.1	53.0	36.8	10.2
	Hispanic	2752	30.7	8.7	0.88	82.7	34.6	50.8	41.3	8.0
	Asian	442	35.6	9.9	0.92	106.2	44.3	27.4	43.0	29.6
	White	15300	36.7	7.9	0.88	108.8	36.1	22.8	50.8	26.5
Special Ed	No	17793	36.6	7.8	0.87	107.8	35.4	23.5	51.5	25.0
	Yes	2607	26.4	8.8	0.87	66.4	33.4	70.6	25.9	3.6
ELL	No	19700	35.6	8.4	0.89	104.0	37.2	27.9	49.2	23.0
	Yes	700	25.2	8.4	0.86	61.8	31.1	77.1	20.9	2.0
FLS	No	11363	37.9	7.4	0.87	114.0	35.0	18.1	51.3	30.6
	Yes	9037	32.0	8.9	0.89	88.2	36.3	43.9	44.3	11.8

Grade 11

Group	Subgroup	Valid N	Raw Scores		Alpha	Scale Scores		Percent in Performance Level		
			Mean	SD		Mean	SD	Basic	Proficient	Advanced
Overall		20542	34.7	8.4	0.89	101.0	39.7	31.4	50.4	18.2
Gender	Male	10403	33.7	8.8	0.89	96.8	40.6	35.6	48.1	16.3
	Female	10139	35.6	7.9	0.88	105.4	38.2	27.1	52.7	20.2
Ethnicity	African American	1324	28.7	9.3	0.89	74.3	39.6	57.9	36.0	6.2
	American Indian	292	29.4	8.7	0.88	76.9	36.7	55.1	39.7	5.1
	Hispanic	2276	30.3	8.5	0.87	80.7	36.9	52.5	40.7	6.8
	Asian	424	34.6	9.0	0.90	101.5	42.5	32.5	46.0	21.5
	White	16226	35.9	7.8	0.87	106.5	38.0	25.8	53.2	20.9
Special Ed	No	18307	35.7	7.8	0.87	105.7	37.6	26.3	53.7	20.1
	Yes	2235	26.0	8.5	0.86	62.6	35.0	73.5	23.4	3.1
ELL	No	20050	34.9	8.3	0.88	102.2	39.1	30.1	51.2	18.6
	Yes	492	23.7	7.5	0.82	53.2	29.9	83.9	14.8	1.2
FLS	No	13234	36.4	7.7	0.87	109.1	37.8	23.5	53.4	23.1
	Yes	7308	31.5	8.8	0.88	86.4	38.6	45.7	45.0	9.3

Appendix G: Contrasting Groups Analyses

Table G.1: Contrasting Group Detail for Grade 3

Raw Score	Teacher Rating						
	Below	Meets	Exceeds	Total	Likelihood ofi Basic	Likelihood of Prof	Logit Ability
11	5	2	0	7	0.71	1.00	-3
12	8	0	0	8	1.00	1.00	-2
13	9	0	0	9	1.00	1.00	-2
14	11	1	0	12	0.92	1.00	-2
15	14	1	0	15	0.93	1.00	-2
16	13	0	0	13	1.00	1.00	-2
17	14	1	0	15	0.93	1.00	-2
18	13	1	0	14	0.93	1.00	-2
19	19	1	0	20	0.95	1.00	-2
20	25	4	0	29	0.86	1.00	-2
21	18	3	0	21	0.86	1.00	-1
22	15	5	0	20	0.75	1.00	-1
23	21	6	0	27	0.78	1.00	-1
24	24	2	0	26	0.92	1.00	-1
25	23	12	0	35	0.66	1.00	-1
26	20	12	1	33	0.61	0.92	-1
27	18	10	0	28	0.64	1.00	-1
28	10	14	1	25	0.40	0.93	-1
29	27	23	0	50	0.54	1.00	-1
30	19	29	0	48	0.40	1.00	-1
31	18	11	3	32	0.56	0.79	0
32	27	27	8	62	0.44	0.77	0
33	31	31	2	64	0.48	0.94	0
34	19	32	2	53	0.36	0.94	0
35	18	44	6	68	0.26	0.88	0
36	17	33	8	58	0.29	0.80	0
37	11	47	20	78	0.14	0.70	0
38	18	52	14	84	0.21	0.79	1
39	8	48	13	69	0.12	0.79	1
40	8	56	21	85	0.09	0.73	1
41	5	49	35	89	0.06	0.58	1
42	0	34	39	73	0.00	0.47	2
43	5	32	36	73	0.07	0.47	2
44	1	16	23	40	0.03	0.41	3
45	0	5	27	32	0.00	0.16	4
	522	646	259	1427			
Mean Logit	-0.833	0.389	1.499	0.143			
SD of Logit	0.953	1.007	1.150	1.319			
SE	0.052	0.050	0.089	0.044			

Table G.2: Contrasting Group Detail for Grade 4

Raw Score	Teacher Rating			Total	Likeli Basic	Likeli Prof	Logit Ability
	Basic	Prof	Adv				
0	1	0	0	1	1.00	1.00	-6.675
8	0	1	0	1	0.00	1.00	-3.020
11	4	1	0	5	0.80	1.00	-2.573
12	7	0	0	7	1.00	1.00	-2.443
13	9	0	0	9	1.00	1.00	-2.319
14	11	0	0	11	1.00	1.00	-2.201
15	12	1	0	13	0.92	1.00	-2.087
16	11	1	1	13	0.85	0.50	-1.976
17	12	2	0	14	0.86	1.00	-1.868
18	15	2	0	17	0.88	1.00	-1.762
19	21	3	0	24	0.88	1.00	-1.658
20	14	5	0	19	0.74	1.00	-1.556
21	19	2	0	21	0.90	1.00	-1.454
22	21	5	1	27	0.78	0.83	-1.352
23	16	6	0	22	0.73	1.00	-1.251
24	19	9	0	28	0.68	1.00	-1.150
25	18	8	1	27	0.67	0.89	-1.047
26	29	16	0	45	0.64	1.00	-0.944
27	25	17	2	44	0.57	0.89	-0.839
28	19	8	2	29	0.66	0.80	-0.733
29	24	13	1	38	0.63	0.93	-0.624
30	30	32	1	63	0.48	0.97	-0.512
31	26	26	1	53	0.49	0.96	-0.396
32	20	24	4	48	0.42	0.86	-0.276
33	30	41	4	75	0.40	0.91	-0.151
34	14	50	9	73	0.19	0.85	-0.020
35	20	61	14	95	0.21	0.81	0.119
36	16	54	12	82	0.20	0.82	0.268
37	9	58	16	83	0.11	0.78	0.429
38	8	44	22	74	0.11	0.67	0.605
39	10	62	30	102	0.10	0.67	0.801
40	4	54	27	85	0.05	0.67	1.025
41	1	23	34	58	0.02	0.40	1.289
42	3	22	32	57	0.05	0.41	1.618
43	0	9	29	38	0.00	0.24	2.064
44	0	10	26	36	0.00	0.28	2.798
45	0	0	2	2	0.00	0.00	4.030
	498	670	271	1439			
Mean Logit	-0.859	0.105	0.557	0.005			
SD of Logit	0.868	0.851	1.342	1.120			
SE	0.049	0.041	0.102	0.037			

Table G.3: Contrasting Group Detail for Grade 5

Raw Score	Teacher Rating			Total	Likeli Basic	Likeli Prof	Logit Ability
	Basic	Prof	Adv				
3	0	0	1	1	1.00	0.00	-4.338
8	1	0	0	1	1.00	1.00	-3.135
11	3	0	0	3	1.00	1.00	-2.680
12	4	1	0	5	0.80	1.00	-2.546
13	3	1	0	4	0.75	1.00	-2.419
14	4	0	0	4	1.00	1.00	-2.298
15	8	0	0	8	1.00	1.00	-2.181
16	8	1	0	9	0.89	1.00	-2.067
17	13	0	0	13	1.00	1.00	-1.957
18	9	0	0	9	1.00	1.00	-1.850
19	13	0	0	13	1.00	1.00	-1.744
20	9	3	0	12	0.75	1.00	-1.641
21	8	4	0	12	0.67	1.00	-1.539
22	16	0	0	16	1.00	1.00	-1.437
23	10	3	0	13	0.77	1.00	-1.337
24	18	4	0	22	0.82	1.00	-1.237
25	13	3	0	16	0.81	1.00	-1.137
26	17	8	0	25	0.68	1.00	-1.037
27	14	6	0	20	0.70	1.00	-0.936
28	17	7	1	25	0.68	0.88	-0.834
29	20	11	2	33	0.61	0.85	-0.732
30	25	17	1	43	0.58	0.94	-0.627
31	16	17	2	35	0.46	0.89	-0.521
32	19	23	3	45	0.42	0.88	-0.412
33	16	29	3	48	0.33	0.91	-0.301
34	18	28	5	51	0.35	0.85	-0.185
35	12	22	6	40	0.30	0.79	-0.066
36	11	38	7	56	0.20	0.84	0.059
37	15	38	13	66	0.23	0.75	0.190
38	11	35	21	67	0.16	0.63	0.329
39	6	49	12	67	0.09	0.80	0.478
40	4	39	20	63	0.06	0.66	0.638
41	0	30	20	50	0.00	0.60	0.814
42	1	27	25	53	0.02	0.52	1.010
43	2	20	34	56	0.04	0.37	1.234
44	2	12	18	32	0.06	0.40	1.499
45	0	4	22	26	0.00	0.15	1.828
46	0	3	20	23	0.00	0.13	2.276
47	0	3	6	9	0.00	0.33	3.010
48	0	0	3	3	1.00	0.00	4.244
	366	486	245	1097			
Mean Logit	-0.870	0.173	0.994	0.009			
SD of Logit	0.804	0.748	0.929	1.069			
SE	0.053	0.042	0.074	0.040			

Table G.4: Contrasting Group Detail for Grade 6

Raw Score	Teacher Rating			Total	Likeli Basic	Likeli Prof	Logit Ability
	Basic	Prof	Adv				
0	1	0	0	1	1.00	1.00	-6.750
5	0	0	1	1	1.00	0.00	-3.736
9	1	0	0	1	1.00	1.00	-2.983
10	4	0	0	4	1.00	1.00	-2.836
12	3	0	0	3	1.00	1.00	-2.571
13	4	0	0	4	1.00	1.00	-2.449
14	4	0	0	4	1.00	1.00	-2.332
15	4	1	0	5	0.80	1.00	-2.220
16	10	0	0	10	1.00	1.00	-2.112
17	2	0	0	2	1.00	1.00	-2.007
18	11	0	0	11	1.00	1.00	-1.905
19	7	1	0	8	0.88	1.00	-1.805
20	6	1	0	7	0.86	1.00	-1.706
21	11	3	0	14	0.79	1.00	-1.609
22	17	3	0	20	0.85	1.00	-1.514
23	21	6	0	27	0.78	1.00	-1.418
24	11	5	0	16	0.69	1.00	-1.323
25	15	6	0	21	0.71	1.00	-1.229
26	18	6	0	24	0.75	1.00	-1.134
27	13	8	0	21	0.62	1.00	-1.038
28	10	10	0	20	0.50	1.00	-0.942
29	17	9	0	26	0.65	1.00	-0.844
30	11	15	1	27	0.41	0.94	-0.745
31	14	15	1	30	0.47	0.94	-0.644
32	24	17	1	42	0.57	0.94	-0.540
33	12	20	0	32	0.38	1.00	-0.433
34	12	35	2	49	0.24	0.95	-0.323
35	14	24	5	43	0.33	0.83	-0.208
36	11	33	4	48	0.23	0.89	-0.088
37	13	38	14	65	0.20	0.73	0.038
38	7	37	10	54	0.13	0.79	0.172
39	8	53	18	79	0.10	0.75	0.316
40	5	38	16	59	0.08	0.70	0.472
41	4	44	30	78	0.05	0.59	0.643
42	6	41	34	81	0.07	0.55	0.834
43	4	38	36	78	0.05	0.51	1.053
44	2	27	52	81	0.02	0.34	1.313
45	1	18	28	47	0.02	0.39	1.638
46	0	7	25	32	0.00	0.22	2.080
47	0	2	19	21	0.00	0.10	2.809
48	1	0	5	6	0.00	0.00	4.039
	339	561	302	1202			
Mean Logit	-0.905	0.183	1.119	0.117			
SD of Logit	0.933	0.781	0.862	1.124			
SE	0.063	0.041	0.062	0.041			

Table G.5: Contrasting Group Detail for Grade 7

Raw Score	Teacher Rating			Total	Likeli Basic	Likeli Prof	Logit Ability
	Basic	Prof	Adv				
0	1	0	0	1	1.00	1.00	-6.526
5	1	0	0	1	1.00	1.00	-3.500
7	1	0	0	1	1.00	1.00	-3.076
8	1	0	0	1	1.00	1.00	-2.899
11	1	0	0	1	1.00	1.00	-2.450
12	4	0	0	4	1.00	1.00	-2.319
13	2	2	0	4	0.50	1.00	-2.194
14	5	0	0	5	1.00	1.00	-2.076
15	7	0	0	7	1.00	1.00	-1.962
16	6	1	0	7	0.86	1.00	-1.851
17	11	0	0	11	1.00	1.00	-1.744
18	10	0	0	10	1.00	1.00	-1.640
19	14	3	0	17	0.82	1.00	-1.538
20	14	0	0	14	1.00	1.00	-1.438
21	8	1	0	9	0.89	1.00	-1.339
22	8	2	0	10	0.80	1.00	-1.241
23	14	8	0	22	0.64	1.00	-1.144
24	16	6	0	22	0.73	1.00	-1.048
25	20	3	1	24	0.83	0.75	-0.952
26	18	9	1	28	0.64	0.90	-0.855
27	15	16	0	31	0.48	1.00	-0.758
28	14	21	1	36	0.39	0.95	-0.661
29	12	12	0	24	0.50	1.00	-0.562
30	23	13	2	38	0.61	0.87	-0.462
31	13	13	2	28	0.46	0.87	-0.360
32	12	21	1	34	0.35	0.95	-0.256
33	9	27	5	41	0.22	0.84	-0.148
34	15	27	4	46	0.33	0.87	-0.038
35	7	21	6	34	0.21	0.78	0.077
36	6	27	7	40	0.15	0.79	0.197
37	5	30	7	42	0.12	0.81	0.323
38	5	27	9	41	0.12	0.75	0.457
39	3	31	25	59	0.05	0.55	0.600
40	4	27	25	56	0.07	0.52	0.755
41	2	26	22	50	0.04	0.54	0.924
42	0	20	32	52	0.00	0.38	1.114
43	0	16	30	46	0.00	0.35	1.332
44	0	12	25	37	0.00	0.32	1.590
45	1	11	24	36	0.03	0.31	1.912
46	0	2	9	11	0.00	0.18	2.351
47	0	0	10	10	0.00	0.00	3.077
48	0	0	3	3	0.00	0.00	4.305
	308	435	251	994			
Mean Logit	-0.869	0.197	1.128	0.104			
SD of Logit	0.806	0.769	0.830	1.093			
SE	0.057	0.046	0.065	0.043			

Table G.6: Contrasting Group Detail for Grade 8

Raw Score	Teacher Rating			Total	Likeli Basic	Likeli Prof	Logit Ability
	Basic	Prof	Adv				
0	2	0	0	2	1.00	1.00	-6.558
7	0	0	1	1	1.00	0.00	-3.155
8	1	0	0	1	1.00	1.00	-2.984
11	2	0	0	2	1.00	1.00	-2.552
12	3	0	0	3	1.00	1.00	-2.426
13	5	0	0	5	1.00	1.00	-2.307
14	9	0	0	9	1.00	1.00	-2.194
15	3	1	0	4	0.75	1.00	-2.085
16	4	0	0	4	1.00	1.00	-1.980
17	14	2	0	16	0.88	1.00	-1.878
18	10	0	0	10	1.00	1.00	-1.779
19	18	0	0	18	1.00	1.00	-1.682
20	14	1	1	16	0.88	0.50	-1.587
21	13	2	0	15	0.87	1.00	-1.493
22	11	1	0	12	0.92	1.00	-1.401
23	13	3	1	17	0.76	0.75	-1.309
24	15	3	0	18	0.83	1.00	-1.219
25	12	13	1	26	0.46	0.93	-1.128
26	19	8	0	27	0.70	1.00	-1.038
27	14	15	1	30	0.47	0.94	-0.947
28	10	9	2	21	0.48	0.82	-0.856
29	18	10	0	28	0.64	1.00	-0.764
30	17	14	4	35	0.49	0.78	-0.671
31	9	27	1	37	0.24	0.96	-0.577
32	12	30	2	44	0.27	0.94	-0.481
33	15	28	5	48	0.31	0.85	-0.383
34	15	35	3	53	0.28	0.92	-0.283
35	16	29	8	53	0.30	0.78	-0.179
36	8	39	4	51	0.16	0.91	-0.072
37	4	34	12	50	0.08	0.74	0.040
38	2	34	10	46	0.04	0.77	0.157
39	7	38	15	60	0.12	0.72	0.280
40	5	46	32	83	0.06	0.59	0.411
41	3	44	30	77	0.04	0.59	0.551
42	0	41	27	68	0.00	0.60	0.704
43	2	19	34	55	0.04	0.36	0.871
44	0	21	40	61	0.00	0.34	1.059
45	1	10	47	58	0.02	0.18	1.274
46	0	9	22	31	0.00	0.29	1.530
47	0	4	26	30	0.00	0.13	1.850
48	0	3	15	18	0.00	0.17	2.288
49	0	3	16	19	0.00	0.16	3.013
50	0	0	3	3	0.00	0.00	4.239
	326	576	363	1265			
Mean Logit	-1.031	0.066	0.959	0.037			
SD of Logit	0.839	0.705	0.883	1.081			
SE	0.058	0.037	0.058	0.038			

Table G.7: Contrasting Group Detail for Grade 11

Raw Score	Teacher Rating			Total	Likeli Basic	Likeli Prof	Logit Ability
	Basic	Prof	Adv				
5	0	1	0	1	0.00	1.00	-3.576
11	1	0	0	1	1.00	1.00	-2.551
12	0	1	0	1	0.00	1.00	-2.422
13	3	0	0	3	1.00	1.00	-2.301
14	7	1	0	8	0.88	1.00	-2.185
15	4	0	0	4	1.00	1.00	-2.073
16	8	0	0	8	1.00	1.00	-1.965
17	9	1	0	10	0.90	1.00	-1.860
18	10	1	0	11	0.91	1.00	-1.757
19	10	0	0	10	1.00	1.00	-1.657
20	5	3	0	8	0.63	1.00	-1.559
21	21	1	0	22	0.95	1.00	-1.462
22	11	4	0	15	0.73	1.00	-1.366
23	19	5	0	24	0.79	1.00	-1.272
24	12	10	0	22	0.55	1.00	-1.177
25	18	8	0	26	0.69	1.00	-1.083
26	17	8	1	26	0.65	0.89	-0.989
27	20	10	0	30	0.67	1.00	-0.895
28	20	19	0	39	0.51	1.00	-0.800
29	17	13	1	31	0.55	0.93	-0.705
30	20	16	2	38	0.53	0.89	-0.608
31	15	21	3	39	0.38	0.88	-0.510
32	18	19	3	40	0.45	0.86	-0.410
33	12	37	6	55	0.22	0.86	-0.308
34	10	23	7	40	0.25	0.77	-0.204
35	10	39	9	58	0.17	0.81	-0.096
36	11	39	9	59	0.19	0.81	0.016
37	21	44	12	77	0.27	0.79	0.132
38	8	44	21	73	0.11	0.68	0.253
39	4	43	28	75	0.05	0.61	0.381
40	3	54	23	80	0.04	0.70	0.516
41	4	47	27	78	0.05	0.64	0.662
42	8	49	35	92	0.09	0.58	0.819
43	5	30	43	78	0.06	0.41	0.992
44	2	26	23	51	0.04	0.53	1.185
45	1	25	34	60	0.02	0.42	1.407
46	0	13	30	43	0.00	0.30	1.669
47	0	10	25	35	0.00	0.29	1.996
48	0	7	19	26	0.00	0.27	2.442
49	0	0	9	9	0.00	0.00	3.175
50	0	0	1	1	0.00	0.00	4.407
	364	672	371	1407			
Mean Logit	-0.387	0.110	0.490	0.082			
SD of Logit	1.015	0.805	1.175	1.020			
SE	0.066	0.039	0.076	0.034			

Appendix H: Cut Scores and Impacts by Method

Table 5.3.1: Grade 3

BookMark	Raw	Logit	Impact
Basic			19.5
Proficient	25	-1.0342	35.4
Advanced	36	0.2372	45.1

Contrasting Groups	Raw	Logit	Impact
Basic			35.8
Proficient	31	-0.4053	49.5
Advanced	42	1.5576	14.7

Table 5.3.2: Grade 4

BookMark	Raw	Logit	Impact
Basic			16.8
Proficient	25	-1.0473	44.6
Advanced	37	0.4288	38.6

Contrasting Groups	Raw	Logit	Impact
Basic			34.0
Proficient	31	-0.3961	50.4
Advanced	41	1.2894	15.6

Table 5.3.3: Grade 5

BookMark	Raw	Logit	Impact
Basic			15.4
Proficient	26	-1.0366	49.3
Advanced	39	0.4777	35.3

Contrasting Groups	Raw	Logit	Impact
Basic			29.0
Proficient	31	-0.5209	56.5
Advanced	43	1.2342	14.5

Table 5.3.4: Grade 6

BookMark	Raw	Logit	Impact
Basic			20.6
Proficient	29	-0.8441	42.1
Advanced	40	0.4716	37.3

Contrasting Groups	Raw	Logit	Impact
Basic			25.7
Proficient	31	-0.6436	60.2
Advanced	44	1.3132	14.1

Table 5.3.5: Grade 7

BookMark	Raw	Logit	Impact
Basic			24.9
Proficient	28	-0.6608	38.8
Advanced	38	0.4568	36.3

Contrasting Groups	Raw	Logit	Impact
Basic			34.4
Proficient	31	-0.3599	49.4
Advanced	42	1.1142	16.2

Table 5.3.6: Grade 8

BookMark	Raw	Logit	Impact
Basic			24.0
Proficient	30	-0.6714	38.4
Advanced	40	0.4110	37.6

Contrasting Groups	Raw	Logit	Impact
Basic			19.3
Proficient	28	-0.8560	58.4
Advanced	43	0.8712	22.3

Table 5.3.7: Grade 11

BookMark	Raw	Logit	Impact
Basic			28.3
Proficient	31	-0.5101	37.9
Advanced	40	0.5162	33.8

Contrasting Groups	Raw	Logit	Impact
Basic			28.3
Proficient	31	-0.5101	58.1
Advanced	44	1.1853	13.6

Appendix I: Panelist Evaluation Form

NEBRASKA STATE ACCOUNTABILITY-READING (NESA-R)
STANDARD SETTING MEETING
JUNE 28-30, 2010
EVALUATION FORM

THE PURPOSE OF THIS EVALUATION IS TO OBTAIN YOUR OPINIONS ABOUT THE STANDARD SETTING MEETING. YOUR OPINION WILL PROVIDE A BASIS FOR EVALUATING THE BOOKMARK PROCESS. PLEASE DO NOT PUT YOUR NAME ON THIS FORM. WE WANT YOUR OPINIONS TO REMAIN ANONYMOUS.

1. Grade Level:

3, 4, or 5 6, 7, or 8 High School Higher Education

2. Circle the phrase that most accurately reflects your satisfaction with the training.

Clarity	Not at all	Somewhat	Adequate	Totally clear
Amount of Time	Way too little	Too Little	Appropriate	Too Much
Practice Exercises	Not Useful	Somewhat Useful	Useful	Very Useful

3. Check the column that most accurately reflects your level of agreement regarding the PLDs.

	Strongly Disagree	Disagree	Agree	Strongly Agree
Adequate information was provided to participants regarding the PLDs.				
Adequate time was provided for participants to gain understanding of the PLDs.				
The PLDs capture what students should know and be able to do at each grade level.				
The PLDs communicate a reasonable profile of students' achievement at Basic, Proficient, and Advanced.				
The PLDs were helpful in making decisions regarding cut-points.				

4. Check the column that most accurately reflects your opinion regarding the usefulness of the following materials.

Materials	Not Useful	Partially Useful	Useful	Very Useful
Test Booklet				
Ordered Item Booklet				
Item Separation Chart				
Item Map				
Statistical Impact Data				

5. Check the column that most accurately reflects your opinion regarding the amount of time allotted for your ratings.

Time Allotted	Too Little Time	About Right	Too Much Time
Round 1			
Round 2			
Round 3			

-DRC- 1 of 2

6. Check the column that most accurately reflects your satisfaction with the following roles.

Role	Not Satisfied	Partially Satisfied	Satisfied	Very Satisfied
DRC Psychometric Lead				
DRC Room Facilitator				
Other DRC Staff				

7. Check the column that most accurately reflects the level of confidence you had in determining the bookmark location for each assessment cut-point. Please only indicate confidence level for the grades in which you participated. Otherwise, leave it blank.

Grade	Cut-point Location	Not Confident	Partially Confident	Confident	Very Confident
3	Basic/Proficient				
	Proficient/Advanced				
4	Basic/Proficient				
	Proficient/Advanced				
5	Basic/Proficient				
	Proficient/Advanced				
6	Basic/Proficient				
	Proficient/Advanced				
7	Basic/Proficient				
	Proficient/Advanced				
8	Basic/Proficient				
	Proficient/Advanced				
11	Basic/Proficient				
	Proficient/Advanced				

8. How confident are you that the processes and methods used will produce valid results?

Not Confident Partially Confident Confident Very Confident

9. If you have further comments or suggestions for ways to improve the Bookmark meeting, please do so in the space below. All comments will remain anonymous.

THANK YOU FOR PARTICIPATING IN THE STANDARD SETTING METTING.

-DRC- 2 of 2

Appendix J: Bookmark Panelist Evaluation Summary

	Grade	3	4	5	6	7	8	11
	Count	41	41	41	33	33	61	27
Training	**Clarity**	3.2	3.0	3.5	3.5	2.5	3.4	3.0
	Time allotted	3.1	3.3	2.8	2.9	3.0	3.4	3.4
	Excer	2.8	3.0	3.0	2.8	2.9	2.6	3.1
PLD's	**Adeq info**	3.4	3.3	3.5	3.2	3.4	3.2	3.4
	Adeq time	3.3	3.4	3.5	3.1	3.4	3.2	3.3
	Capture	3.2	3.3	3.4	2.8	3.3	3.1	3.1
	Comm	3.2	3.0	3.4	3.0	3.3	3.0	2.9
	Helpful	3.3	3.3	3.4	2.8	3.2	3.0	3.0
Materials	**Test bklt**	3.6	3.7	3.7	3.7	3.7	3.5	3.6
	OIB	3.6	3.6	3.7	3.6	3.6	3.6	3.6
	Item sep	3.4	3.4	3.7	3.4	3.2	3.4	3.3
	Item map	3.3	3.3	3.5	3.2	3.1	3.2	3.3
	Stat data	3.5	3.3	3.6	3.4	3.1	3.2	3.1
Amount of time*	**Rnd 1**	2.4	2.0	2.0	1.9	2.0	2.5	2.1
	Rnd 2	2.2	2.5	2.0	2.0	2.3	2.4	2.4
	Rnd 3	1.6	2.2	1.9	1.7	2.2	1.9	2.3
Roles	**PS Lead**	3.4	3.1	3.6	3.6	3.1	3.3	3.3
	Rm Fac	3.6	3.1	3.7	3.5	2.2	3.5	3.4
	Other	3.4	3.3	3.6	3.3	3.2	3.4	3.2
Confidence	**Below/Meets**	3.0	3.1	3.4	3.2	3.2	3.0	3.4
	Meets/Exceeds	2.9	2.9	3.4	2.9	2.9	3.0	2.9
Process	**Confid**	3.0	2.7	3.3	3.0	3.0	2.6	3.1

*Three point scale: Too Little, About Right, Too Much

For the quantitative analyses, the categories were coded 1 to 4, except questions about "Amount of Time" were 1 to 3. Please refer to Appendix I for the precise category labels.

Appendix K: Cut Scores and Standard Errors of Measurement by Round

Reading		Round 1		Round 2		Round 3	
Grade	Level	Median	SE of Median	Median	SE of Median	Median	SE of Median
3	Below/Meets	15	0.73	15	0.54	15	0.46
	Meets/Exceeds	36	0.74	37	0.52	41	0.52
4	Below/Meets	12	0.82	11	0.48	11	0.73
	Meets/Exceeds	34	0.88	35	0.57	39	0.58
5	Below/Meets	14	0.65	14	0.46	14	0.30
	Meets/Exceeds	41	0.83	41	0.58	41	0.56
6	Below/Meets	13	0.97	15	0.76	16	0.71
	Meets/Exceeds	41	1.03	44	0.69	44	0.72
7	Below/Meets	14	0.90	12	0.23	14	0.26
	Meets/Exceeds	38	0.89	38	0.15	40	0.52
8	Below/Meets	17	0.77	17	0.50	17	0.51
	Meets/Exceeds	42	0.72	44	0.49	44	0.44
11	Below/Meets	19	1.29	19.5	0.78	20	1.02
	Meets/Exceeds	36.5	0.95	38	0.70	42	0.45