



Technical Report

Nebraska State Accountability (NeSA)

Spring 2013 Writing Test

Grades 4, 8, and 11

July 2013





TABLE OF CONTENTS

GENERAL INFORMATION	1
HISTORY	1
OVERVIEW	2
ADMINISTRATION OF THE WRITING ASSESSMENT	4
WRITING TOPICS	4
TEST SESSIONS, TIMING, AND FORMAT	4
SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS.....	6
MATERIALS RETURN	6
TEST SECURITY MEASURES	6
SAMPLE MANUALS.....	7
PROCESSING AND SCORING THE NESA-WRITING	8
RECEIPT OF MATERIALS	8
SCANNING OF MATERIALS.....	8
MATERIALS STORAGE	8
PERFORMANCE ASSESSMENT SERVICES (PAS).....	10
RANGEFINDING.....	10
TRAINING MATERIAL CREATION	11
READER RECRUITMENT/QUALIFICATIONS	11
TEAM LEADER AND READER TRAINING.....	11
HANDSCORING PROCESS.....	12
QUALITY CONTROL	12
DECISION CONSISTENCY	14
STANDARD SETTING.....	16
REPORTING.....	23
GRADE 4, 8, AND 11 REPORTS.....	23
APPENDIX A: NEBRASKA DEPARTMENT OF EDUCATION SCORING	
GUIDE FOR NARRATIVE WRITING – ANALYTIC – GRADE 4.....	25
APPENDIX B: NEBRASKA DEPARTMENT OF EDUCATION SCORING	
GUIDE FOR DESCRIPTIVE WRITING – ANALYTIC – GRADE 8.....	26
APPENDIX C: NEBRASKA DEPARTMENT OF EDUCATION SCORING	
GUIDE FOR PERSUASIVE WRITING – ANALYTIC – GRADE 11	27
APPENDIX D: PERFORMANCE LEVEL DESCRIPTORS GRADE 4	28
APPENDIX E: PERFORMANCE LEVEL DESCRIPTORS GRADE 8	29
APPENDIX F: PERFORMANCE LEVEL DESCRIPTORS GRADE 11	30
APPENDIX G: STANDARD SETTING PANELIST EVALUATION FORM	31

APPENDIX H: COMPOSITE TO SCALE SCORE TABLES GRADE 4	33
APPENDIX I: COMPOSITE TO SCALE SCORE TABLES GRADE 8.....	35
APPENDIX J: COMPOSITE TO SCALE SCORE TABLES GRADE 11	37
APPENDIX K: FORMATTING INFORMATION FOR TAC	39
APPENDIX L : PERFORMANCE ASSESSMENT SERVICES FORMATTING RESPONSE.....	46

GENERAL INFORMATION

HISTORY

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 (NeSA-Reading, NeSA-Mathematics, and NeSA-Science) under the direction of the Department of Education. Legislative Bill (LB) 1157 passed by the 2008 Nebraska Legislature (<http://www.legislature.ne.gov/laws/statutes.php?statute=79-760.03>) requires a single statewide assessment of writing, reading, mathematics, and science in Nebraska's K-12 public schools against the Nebraska academic content standards.

The legislation requires that:

- The assessments will be used for accountability purposes.
- The assessments will be criterion-referenced.

The NDE prescribed such assessments starting in the 2009-2010 school year and phased in as described in Table 1-1. The state uses the expertise and experience of the educators in the state to participate, to the maximum extent possible, in the design and development of the statewide assessment system.

Table 1-1 NeSA Administration Schedule

Subject	Administration Year		Grades
	Field Test	Operational	
Reading	2009	2010	3 through 8 plus 1 high school
Mathematics	2010	2011	3 through 8 plus 1 high school
Science	2011	2012	At least 1 grade in elementary, middle/junior high, and high school

In October 2010, the NDE contracted with DRC to provide and operate a computerized information system to support the administration, record keeping, and reporting for the statewide student NeSA-Writing assessment under the direction of the Department of Education.

NeSA-Writing will be phased in as described in Table 1-2.

Table 1-2 NeSA-Writing Administration Schedule

Year	Paper/Pencil Mode	Online Mode
2011	Grades 4 and 8	Grade 11, Pilot Year
2012	Grade 4	Grades 8 and 11
2013	Grade 4	Grades 8 and 11

A governor-appointed Technical Advisory Committee (TAC) consisting of three nationally recognized experts in assessment and measurements, one local administrator, and one teacher from Nebraska provides technical advice, guidance, and research to help NDE make informed decisions regarding standards, assessment, and accountability.

OVERVIEW

The NeSA tests are developed specifically for Nebraska. Since 2002, the Nebraska statewide writing assessment has been annually administered in grades 4, 8, and 11 for the purpose of providing school districts with instructional information and to include writing results from grades 4 and 8 as the “other academic indicator” in the federal accountability requirements of the Elementary and Secondary Education Act (ESEA).

The Nebraska statewide writing assessment is intended to:

1. Gather information to assist teachers in determining the progress of students in meeting state or local standards for writing;
2. Provide each local school district with a report of student progress in meeting state or local standards for writing; and
3. Lead to improved writing by Nebraska students.

DRC and Computerized Assessments and Learning (CAL) were the providers of the printed and online versions, respectively, of the 2013 NeSA-Writing Tests.

Paper/Pencil and Online Testing Window: January 21 – February 8, 2013

Number of Potential Testing Sites

254 districts

949 schools

Background Information Regarding the Formatting Issues with the Online NeSA-Writing Engine at Grades 8 and 11

- Some students taking NeSA-Writing at Grades 8 and 11 online experienced formatting issues.
Examples:
 - Words breaking at the end of lines
 - Odd wrap-arounds
 - Centering could not be turned off
- The issues initially appeared to be random and infrequent.
- The contractor was unable to fix all the problems during the test window. In addition, NDE determined that releasing an updated engine during the writing window would create an inequitable test administration across districts in Nebraska.
- The formatting problems did not cause any student's work to be lost. The formatting problems do not appear to have significantly lowered any scores.
 - The state averages are up slightly from last year.
 - Students whose papers appeared to have been affected scored higher than those unaffected.

- The Sentence Fluency/Conventions domain scores were not lower than the other domains.

Release Decisions

- Release percent of number and percent of students at each performance level: Below, Meets, Exceeds and release Average Scale Scores for schools, for districts, and for the State of Nebraska. Release same disaggregated information—all asterisked with the following information:
 - *Students at grades 8 and 11 experienced formatting issues with the NeSA-Writing online test administration. While research into the score results does not indicate an effect on student results, it also does not assure there was no effect. Scores should be interpreted with caution.*
- NeSA-Writing status scores or improvement scores will not be included in Nebraska State Accountability System [NePAS] in writing at grades 8 and 11.
- Individual Student Reports will include the student's individual scale score and performance level, but will also include the same italicized information as above.
- The problems will be fixed before the next testing period. A new test engine is being used to administer all online tests—INSIGHT.

ADMINISTRATION OF THE WRITING ASSESSMENT

WRITING TOPICS

At each grade level, students responded to a writing topic developed by NDE to measure composition of writing as specified in the writing content standards. Each student responded to one writing topic in a specific mode. The types of the writing topics for each grade were as follows:

- Grade 4 – Narrative
- Grade 8 – Descriptive
- Grade 11 – Persuasive

TEST SESSIONS, TIMING, AND FORMAT

The test window for the grade 4 paper/pencil tests, including make-up tests, was January 21 – February 8, 2013. The grade 4 tests were administered in two independent sessions on two consecutive days. Each session was 40 minutes, unless a student's IEP or 504 Plan called for additional time. Spanish versions of these tests were developed and made available by DRC for any district that requested them. All student responses were returned to DRC using standard writing booklets for processing and scoring.

The test window for the grades 8 and 11 tests, including make-up tests, was January 21 – February 8, 2013. The majority of students were administered the test online in one session. Students were allowed to use paper to pre-write and continued their work online by drafting and finalizing their response. It was recommended by NDE that districts schedule 90 minutes for students to complete the assessment; however, the test was not timed, and students were allowed as much time as necessary to complete and submit their final essays. Students with an IEP or 504 Plan were allowed to use a paper/pencil test as an accommodation.

The required grade 4 NeSA-Writing paper/pencil test as well as the grades 8 and 11 NeSA-Writing online tests were available to all schools. Spanish versions of the tests were made available to all districts. Table 2-1 shows the number of student who took each exam by mode of administration.

Table 2-1 2013 NeSA-Writing Test Participation

Grade	Number of Students Tested Paper/Pencil	Number of Students Tested Online
4	22,238	N/A
8	454	20,650
11	446	20,529

Tables 2-2 and 2-3 depict the N count as well as the percentage of students that completed their online test in each time span. Student time span is based on the student's initial login and final log out. Students' tests may be reactivated to allow testing across longer periods of time, even multiple days. Thus, in some cases, the elapsed time may not reflect the actual amount of time a student spent completing the test.

Table 2-2 2013 NeSA-Writing Grade 8 Online Test Times

Time Span in Minutes	Student Count	% in Each Time Span
0-10	79	0.39%
10-20	212	1.05%
20-30	460	2.28%
30-40	996	4.93%
40-50	1774	8.78%
50-60	2603	12.89%
60-70	3087	15.29%
70-80	3085	15.28%
80-90	2517	12.46%
90+	5383	26.65%
Total	20,196	100.00%

Table 2-3 2013 NeSA-Writing Grade 11 Online Test Times

Time Span in Minutes	Student Count	% in Each Time Span
0-10	48	0.24%
10-20	296	1.46%
20-30	892	4.39%
30-40	2201	10.84%
40-50	3273	16.12%
50-60	3504	17.25%
60-70	3232	15.91%
70-80	2565	12.63%
80-90	1629	8.02%
90+	2670	13.15%
Total	20,310	100.00%

SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS

A single shipment was sent out by DRC to each district. The shipment was delivered by January 7, 2013. The shipment contained all necessary materials to complete the NeSA-Writing test administration.

- *Writing Manual for Test Coordinators and Administrators*
- Secure Materials: Standard Writing Booklets and Spanish Translation Booklets (Grades 4, 8, and 11)
- Administrative Materials: Student PreID Labels, District/School Labels, Do Not Score Labels, Return Shipping Labels, etc.

DRC ensured that all assessment materials were assembled correctly prior to shipping. DRC Operations staff used the automated Operations Materials Management System (OpsMMS) to assign secure materials to a district at the time of ship out. This system used barcode technology to provide an automated quality check between items requested for and items shipped to each site. A shipment box manifest was produced and placed in each box shipped. DRC Operations staff double-checked all box contents against the manifest prior to the box being sealed for shipment to ensure accurate delivery of materials. Districts and schools were selected at random and examined for correct and complete packaging and labeling.

OpsMMS, along with the UPS tracking system, allowed DRC to track the items from the point of shipment from DRC's warehouse facility to receipt at the district. All DRC shipping facilities, materials processing facilities, and storage facilities are secure. Access is restricted by security code. Only DRC inventory control personnel have access to stored secure materials. DRC employees are trained in and made aware of the high level of security that is required.

The paper/pencil assessments for grades 4, 8, and 11 were packaged by school, and shipped to districts to the attention of the District Assessment Contacts. DRC packed 32,350 standard writing booklets, 376 Spanish translation booklets, 3,095 manuals, and approximately 4,790 non-secure materials for testing sites. DRC used UPS to deliver materials to the testing sites.

MATERIALS RETURN

The materials return window was February 13-15, 2013. DRC used UPS for all return shipments.

TEST SECURITY MEASURES

Test security is essential to obtaining reliable and valid scores for accountability purposes. The 2013 NeSA-Writing included a Test Security Agreement that was provided to all districts by NDE in Nebraska's *Standards, Assessment, and Accountability Updates*. The agreement was to be signed by every school principal and District Assessment Contact and faxed to NDE by January 18, 2013. The purpose of the agreement was to serve as a tool to document that the individuals responsible for administering the assessments both understood and acknowledged the importance of test security. The Test Security

Agreement attested that all security measures were followed concerning the handling of secure materials.

SAMPLE MANUALS

Copies of the *Writing Manual for Test Coordinators and Administrators* and the *Online Test Administration Manual* can be found on the Nebraska Department of Education website at www.education.ne.gov/assessment.

PROCESSING AND SCORING THE NeSA-WRITING

RECEIPT OF MATERIALS

Receipt of NeSA-Writing materials began on February 13, 2013, and concluded on February 27, 2013. Any materials received after February 27, 2013, were considered late and were checked-in, scanned, and processed during the late window of March 1, 2013 through April 2, 2013. OpsMMS was utilized to receive materials securely, accurately, and efficiently. This system features advanced automation and cutting-edge barcode scanners. Captured data were organized into reports, which provided timely information with respect to suspected missing materials.

The check-in process occurred immediately upon receipt of materials; therefore, DRC provided immediate feedback to districts regarding any missing materials based on actual receipts versus expected receipts. DRC produced and submitted to NDE a Missing Materials Report that listed all standard and Spanish translation writing booklets by district, school, and grade that were not returned to DRC.

SCANNING OF MATERIALS

DRC used its image scanning system to capture student essays. The images were then loaded into the image scoring system for both the hand scoring of student responses, and for the capture of demographic data.

Customized scanning programs for all scannable documents were prepared to read the writing documents and to electronically format the scanned information. Before materials arrived, all image scanning programs went through a quality review process that included scanning of mock data from production booklets to ensure proper data collection.

After each batch of writing booklets was scanned, writing documents were processed through a computer-based edit program to detect potential errors as a result of smudges, multiple marks, and omits in predetermined fields. Marks that did not meet the pre-defined editing standards were routed to human editors for resolution.

Before batches of writing responses were extracted for scoring, a final edit was performed to ensure that all requirements for final processing were met. If a batch contained errors, it was flagged for further review before being extracted for scoring and reporting.

MATERIALS STORAGE

Upon completion of processing, student writing booklets were boxed for security purposes and final storage.

- Project-specific box labels were created containing unique customer and project information, material type, batch number, pallet/box number, and the number of boxes for a given batch.
- Boxes were stacked on project-specific pallets that were labeled with a list of its contents and delivered to the Materials Distribution Center for final secure storage.

- All paper/pencil writing booklets will be securely stored for one year until DRC receives written authorization from NDE requesting that they be permanently destroyed.
- All electronic student response images will be securely stored until DRC receives written authorization from NDE requesting that they be permanently deleted.

PERFORMANCE ASSESSMENT SERVICES

In 2013, NDE continued the use of analytic scoring rubrics for grades 8 and 11 and adopted the use of an analytic scoring rubric for grade 4. These rubrics use a 1-4 scale across four domains to define narrative, descriptive, and persuasive writing performance analytically. The rubrics define qualities of each score point for each of the four domains; Ideas/Content, Organization, Voice/Word Choice, and Sentence Fluency/Conventions.

RANGEFINDING

After receiving student responses from the 2012 NeSA-W Field Test, DRC's Performance Assessment Services (PAS) staff reviewed all of the responses and assembled them into sets that exemplified the range of different score points, for each of the four domains, for each of the three prompts. Copies of these sets were made for each member of the rangefinding committees. DRC's PAS staff then travelled to Lincoln, Nebraska (June 28 and 29, 2012) and facilitated the rangefinding sessions. The rangefinding committees consisted of Nebraska educators, NDE staff members and DRC Performance Assessment Staff. The rangefinding meeting began in a joint session with a review of the history of the assessment and a discussion of the rangefinding process, along with guidelines for the consensus scoring of the assembled responses. The group then broke into three grade specific committees consisting of ten or twelve NE educators, an NDE representative and two DRC facilitators on each committee. Each committee reviewed the current prompt and scoring rubric, and the grade 8 and grade 11 committees also reviewed the 2012 Scoring Guide anchor papers.

Initially, each student response was read aloud and then discussed by all members of the group equally; to ensure that everyone was interpreting the analytic rubric consistently and uniformly. Each of the four domain scores were addressed independently and following the discussions, scores were agreed upon in each domain. The first set of 20 responses was discussed at length and then consensus scored using this method. Committee members then went on to score additional responses independently. For each student response, committee members' scores were recorded and, if needed, were discussed until a consensus was reached. Responses for which there was a strong agreement among committee members were identified as potential anchor papers to be used in the Scoring Guides for training DRC readers. Each committee consensus scored over 100 responses.

Discussions of student responses included the mandatory use of rubric language. This ensured that the committee members remained focused on the specific requirements of each score point in each domain. DRC PAS staff took notes addressing how and why committees arrived at score point decisions and how each range of scores was defined. This information was used by the scoring directors and team leaders during reader training.

TRAINING MATERIAL CREATION

As part of preparation for the 2013 NeSA-Writing assessment, DRC's PAS staff assembled the committee scored rangefinding responses into sets used for training readers. Responses that the rangefinding committee had a strong consensus and were relevant in terms of the scoring concepts they illustrated were annotated and included as anchor papers in a scoring guide. The full range of each score point in each domain was clearly represented and annotated in the Scoring Guide. These anchor papers, along with the grade specific analytic rubric, served as the readers' constant reference throughout the project. Training and qualifying sets were assembled using the student responses that were reviewed and scored by rangefinding committee members. Responses were selected for training to show readers the ranges for each score point in each domain and to highlight some of the writing characteristics within each domain.

Validity papers were selected from current operational student responses, and consensus scored by DRC PAS staff and NDE representatives. These papers were entered into the imaging system in preparation for being scored by all readers. These pre-scored responses were dealt out intermittently to all readers throughout the project as a quality control process. The readers were unaware that these responses served as validity papers with the objective of ensuring that readers scored student responses in a manner consistent with their training and with Nebraska statewide standards throughout the duration of the project.

READER RECRUITMENT/QUALIFICATIONS

DRC retains a pool of experienced readers from year to year and all of the 2013 NeSA-Writing readers came from this population. Every reader had at least one year of previous scoring experience with Nebraska writing.

The Scoring Director and Team Leaders were chosen by the content specialists from a pool, consisting of experienced individuals who are proven successful readers and leaders, and who had strong backgrounds in writing. Those selected demonstrated organization, leadership, and management skills. All scoring personnel were required to sign confidentiality agreements before any training or handling of secure materials began.

TEAM LEADER AND READER TRAINING

Representatives from NDE travelled to the DRC Plymouth, Minnesota Scoring Center (February 7- 13, 2013) to collaborate with DRC Scoring Directors and Team Leaders during a three-day training session. The content specialist, scoring director and representative from NDE worked cooperatively to review and discuss all of the training materials, and to consensus score a number of additional validity papers. Team leaders were required to annotate all of their training materials with notes from the training sessions. To facilitate scoring consistency, it was imperative that each team leader imparted the same rationale for each response as the other team leaders used.

Two days of reader training took place on February 11-12, 2013 for grades 8 and 11, and February 14-15, 2013 for grade 4, at the DRC Scoring Center. Reader training began with the scoring director providing an intensive review of the analytic scoring rubric, and the anchor papers in the scoring guide. Next, readers practiced by independently scoring the

responses in the training sets. After each training set, the scoring director or team leaders led a thorough discussion of the responses, either in a room-wide or small-group setting. Once the scoring rubric, anchor sets, and training sets were thoroughly discussed, each rater was required to demonstrate understanding of the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the true scores) on at least one of the qualifying sets. Readers who failed to achieve 70% exact agreement on the first qualifying set were given additional, individual training. Readers who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score any student responses. These individuals were removed from the pool of potential readers in DRC's imaging system and released from the project. 40 readers were qualified to score Nebraska grade 4 student writing responses, 34 readers were qualified to score Nebraska grade 8 student writing responses, and 35 readers were qualified to score Nebraska grade 11 student writing responses.

Following training and qualifying, a period of paired scoring took place, when readers were required to work cooperatively to score live responses and discuss and agree on the appropriate score. Once team leaders were satisfied with their performance, the readers were permitted to score independently while being monitored closely.

HANDSCORING PROCESS

Student responses were scored blindly and independently by multiple readers using DRC's handscoring system. Readers were not able to see demographic information pertaining to the student being scored, nor were they able to see any of the other scores given by any other reader. Each reader was required to apply the analytic scoring rubric to a given writing response and was instructed to avoid any bias in their scoring decisions. Each student paper was scored twice and non-adjacent scores were adjudicated. Data collected from the multiple reads was used to calculate the rater agreement rates and score point distributions. Student responses that were considered non-scoreable (Blank, Refusal, Off-Topic, Foreign Language, Illegible/Incoherent, Insufficient, Copy of Prompt), were automatically routed to the scoring director for review, and then to a content specialist for final approval. Those foreign language papers that were identified as being written in Spanish were then scored by a select group of qualified readers and team leaders who are DRC's specialist Spanish scorers.

QUALITY CONTROL

Validity sets

NDE approved/scored validity responses that were added into the Image Handscoring System for daily quality control checks. These pre-scored responses helped to track consistency over time, and how well individual readers were performing.

Recalibration Tests

During the course of scoring, two recalibration sets were produced using pre-determined scored student responses, and administered to readers as a way to address any scoring issues, and as a method of reinforcing the Nebraska scoring standards set out in the rubric.

Monitoring and Read-Behinds

Team leaders conducted routine read-behinds for every member of their teams and provided feedback and assistance to their readers.

Statistical Handscoring Reports

Numerous quality control reports were produced on demand or run daily in order to maintain high standards of scoring accuracy. The Reader Monitor Report and Score Point Distribution Report were especially helpful in analyzing scoring data and maintaining high standards of scoring quality.

Table 4-1 Reader Agreement rates for NeSA-W 2013

GRADE	IDEAS/CONTENT			ORGANIZATION			VOICE/WORD CHOICE			SENTENCE FLUENCY/CONVENTIONS		
	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ
4	75%	24%	99%	74%	26%	100%	73%	27%	100%	74%	26%	100%
8	77%	23%	100%	77%	23%	100%	76%	24%	100%	75%	25%	100%
11	74%	25%	99%	77%	23%	100%	74%	26%	100%	75%	25%	100%

Table 4-2 Score Point Distributions for NeSA-W 2013

GRADE	IDEAS/CONTENT				ORGANIZATION				VOICE/WORD CHOICE				SENTENCE FLUENCY/CONVENTIONS			
Score Points	%1	%2	%3	%4	%1	%2	%3	%4	%1	%2	%3	%4	%1	%2	%3	%4
4	3	32	54	8	4	33	53	7	4	30	53	10	4	29	53	11
8	1	25	60	14	2	26	61	11	2	23	59	16	2	27	58	12
11	2	18	62	17	3	14	69	13	2	16	64	17	4	18	63	14

Table 4-3 Validity Agreement for NeSA-W 2013

GRADE	IDEAS/CONTENT			ORGANIZATION			VOICE/WORD CHOICE			SENTENCE FLUENCY/CONVENTIONS		
VALIDITY	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ
4	83%	17%	100%	77%	22%	99%	78%	22%	100%	80%	19%	99%
8	88%	11%	99%	85%	15%	100%	84%	16%	100%	79%	21%	100%
11	80%	20%	100%	79%	20%	99%	82%	17%	99%	78%	22%	100%

DECISION CONSISTENCY

In a standards-based testing program, there is great interest in how accurately students are classified into achievement categories. Decision consistency answers the question: What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test (Huynh, 1976). If two equivalent forms were given to the same students, the consistency of the measure would be reflected by the extent that the classification decisions made from the first set of test scores matched the decisions based on the second set of test scores. In contrast to Coefficient Alpha, which describes the relative ordering of students, it is the actual student scores that are important in decision consistency.

Table 4-4 Pseudo-Decision for Two Hypothetical Categories

		TEST ONE		
		LEVEL I	LEVEL II	MARGINAL
TEST TWO	LEVEL I	ϕ_{11}	ϕ_{12}	$\phi_{1\bullet}$
	LEVEL II	ϕ_{21}	ϕ_{22}	$\phi_{2\bullet}$
	MARGINAL	$\phi_{\bullet 1}$	$\phi_{\bullet 2}$	1

Table 4-5 Pseudo-Decision for Four Hypothetical Categories

		TEST ONE				
		LEVEL I	LEVEL II	LEVEL III	LEVEL IV	MARGINAL
TEST TWO	LEVEL I	ϕ_{11}	ϕ_{12}	ϕ_{13}	ϕ_{14}	$\phi_{1\bullet}$
	LEVEL II	ϕ_{21}	ϕ_{22}	ϕ_{23}	ϕ_{24}	$\phi_{2\bullet}$
	LEVEL III	ϕ_{31}	ϕ_{32}	ϕ_{33}	ϕ_{34}	$\phi_{3\bullet}$
	LEVEL IV	ϕ_{41}	ϕ_{42}	ϕ_{43}	ϕ_{44}	$\phi_{4\bullet}$
	MARGINAL	$\phi_{\bullet 1}$	$\phi_{\bullet 2}$	$\phi_{\bullet 3}$	$\phi_{\bullet 4}$	1

If a student is classified as being in one category based on Test One's score, how probable would it be that the student would be classified in the same category based on Test Two?

The proportions of correct decisions, ϕ for two and four categories are computed by the following two formulas, respectively:

$$\phi = \phi_{11} + \phi_{22}$$

$$\phi = \phi_{11} + \phi_{22} + \phi_{33} + \phi_{44}.$$

It is the proportion of students classified by the two forms into exactly the same achievement level that represents the overall consistency.

Since it is not possible to retest in order to estimate the proportion of students who would be reclassified in the same performance levels, a statistical model needs to be imposed on the data in order to project the consistency of classifications solely using data from the available administration (Hambleton & Novick, 1973). Although a number of procedures are available, two well-known methods were developed by Hanson and Brennan (1990) and Livingston and Lewis (1995) utilizing specific True Score Models.

Table 4-6 NeSA-W Decision Consistency Results

Content Area	Grade	Livingston & Lewis				Hanson & Brennan			
		Decision Accuracy		Decision Consistency		Decision Accuracy		Decision Consistency	
		Proficient	Advanced	Proficient	Advanced	Proficient	Advanced	Proficient	Advanced
Writing	4	0.91	0.91	0.87	0.88	0.92	0.92	0.89	0.90
	8	0.91	0.92	0.87	0.89	0.92	0.93	0.88	0.90
	11	0.91	0.91	0.88	0.87	0.92	0.92	0.89	0.89

STANDARD SETTING

Introduction

Academic Performance Levels for the writing component of the Nebraska State Accountability assessments (NeSA-Writing) grades 8 and 11 were developed in Spring 2012 and continued for use in Spring 2013. *Academic Performance Levels* for the writing component of the Nebraska State Accountability assessments (NeSA-Writing) grade 4 were developed in Spring 2013 by establishing cut scores that define operationally the three Performance Levels: *Below the Standards*, *Meets the Standards*, and *Exceeds the Standards*. These Performance Level designations will be used by local, state, and federal accountability programs and are central to communicating to parents, teachers, and the public. Standard setting for grades 8 and 11 was completed in April 2012. As with the previous grades 8 and 11 writing standard setting, grade 4 standard setting process, completed Spring 2013, consisted of three distinct events. First, a meeting was held March 4, 2013 with the Nebraska State Board of Education and other stakeholders to introduce the process and obtain feedback to ensure an effective, defensible process. Second, a *Body of Work* Standard Setting was conducted on March 21, 2013 in Lincoln, Nebraska, after the operational data were available. Finally, recommendations of the *Body of Work* process were presented to the State Board of Education on April 8-9, 2013. The purpose of this meeting was for the State Board of Education to formally establish the Performance Levels for NeSA-Writing grade 4. This report specifically documents the *Body of Work* portion of the process.

Holistic Judgments

A holistic judgment typically requires the appraisal of all the available evidence for each student on the construct of interest. The task is to appraise a unit of work much larger than a test item and determine which of the *Performance Level Descriptors* (PLDs), which define the levels, best describes the student. Non-holistic processes, like *Angoff* and *Bookmark*, require the judges to estimate, by various procedures, the likelihood that a *borderline* candidate will succeed on each item. By definition, the borderline student is on the line between two Performance Levels, but the PLDs describe the typical, not the borderline student in the levels. The description of the borderline student is a negotiated consensus about what is different about two levels. The borderline student should have all or nearly all of the attributes of the lower level and few if any of the higher level. The holistic methods do not require this initial negotiation; it is *simply* a process of matching the student's evidence to a Performance Level.

Body of Work Method

Body of Work (BoW) (Kingston, Kahl, Sweeny & Bay, 2001) has much in common with the *Contrasting Groups* (CG) method. Both require a holistic judgment about the individual and both employ logistic regression to do the arithmetic. With *CG*, the judgment is based on a teacher's direct experience with the student in the classroom; typically garnered just prior to the assessment for which the Performance Levels are being developed. With *BoW*, the judgment is based on a significant sample of the student's work collected during the assessment as direct evidence of proficiency on the construct of interest. The judgment is a holistic evaluation of the evidence without consideration of, or perhaps without knowledge of, the rubric to be used for quantifying the performance.

For the NeSA-W, the construct of interest is writing proficiency. The evidence of a student's proficiency is a prompted writing sample and the task for the judges was to sort the responses into groups corresponding to the Performance Levels defined by the *PLDs*. The responses have been scored, using the established rubrics that are the basis of all reporting and analyses for the NeSA-W. The scores are not explicitly given to the judges and the responses were not arranged in score order. Judges are allowed to place responses into whatever categories they deem appropriate, even if it is not consistent with the scoring.

BoW has five basic steps; two (II and IV) of which involve the judges.

- I. *Selection*: Organizers choose responses that cover the range of possible cut scores.
- II. *Focus*¹: Judges assign sparsely spaced responses to Performance Levels.
- III. *Refinement*: Organizers select new sets of responses clustered near the tentative cut points.
- IV. *Pinpoint*: Judges assign finely spaced responses to Performance Levels.
- V. *Analysis*: Psychometrics computes the final cut point recommendations.

Selection

The initial selection of responses included three or four responses at 20 to 25 score points covering the 70 point weighted score range. Because of the weighting, DRC Psychometric Services staff determined the patterns of domain and reader scores that should be included and DRC PAS staff selected the responses and provided hard copies in order to prepare the judge packets for the Standard Setting meetings.

Responses included were selected to cover the scale score range uniformly. The results of the selection process were used during the Focus and Pinpointing steps. The papers represented the breadth of possible score profiles across domains to provide maximum diversity, so that responses at the same total score arrived at that score by different paths.

For Focusing, the packet given to each judge contained enough responses to cover the maximum possible range of cut scores. It was not necessary for any judge to review more than 15 responses in this step. There was a trade-off between this stage and the Pinpointing step; the finer the spacing at this step, the sharper the focus would be at the Pinpoint step.

For Pinpointing, the packets contained 20 to 25 responses, clustered around the tentative cut points. While the Pinpointing response packets did not include any of the Focusing responses, psychometric calculations can be done such that the results from either round can be combined and made equivalent.

¹ This step is referred to as *Range Finding* in the Standard Setting literature, but to avoid confusion with the hand scoring process, this document will use the term *Focus* to refer to the process of narrowing consideration to scores in the vicinity of the eventual Performance Levels.

Focusing

The purpose of *Focusing* is to narrow the possible range of outcomes for Pinpointing. Each judge was asked to review the responses in the packet and to assign each to a Performance Level. Judges were not shown the scoring rubrics and any who were familiar with the rubrics were cautioned not to attempt to score the responses. The appraisal in this process is a holistic judgment about how the response compares to the PLDs, not to the rubric.

The time-consuming part of this activity is the reading. Because the responses are relatively widely spaced, sorting them into Performance Levels proceeded relatively quickly.

Discussion at the end dealt with the boundaries, where there is a lack of consensus among the judges. Participants were asked to locate their best response below the *Meets the Standards* line and describe, at least to themselves, what prevented it being placed above the line. Then, for the weakest paper above the line, why does it belong there? The discussion then turned to responses just below and just above the *Exceeds the Standards* line.

The process for the judges was:

1. Group discussion of the PLDs.
2. Read the responses in their packets.
3. Assign each response to a Performance Level.
4. Compare the strongest response below the *Meets the Standards* line to the weakest response above the line.
5. Rearrange and reconsider as desired.
6. Group discussion of individual assignments.

Refinement

The analysis required for Refinement is minimal; simply a matter of eliminating regions of score points where there was strong consensus on the appropriate Performance Level. Most of the effort was tabulation of judges' assignments and reorganization of the responses to focus on the areas without consensus. The final performance standard was set at the point of complete disagreement: the score where half the judges place the responses above the line and half, below.

Pinpointing

The task for a judge during Pinpointing is identical to Focusing: review the packet of responses and sorting them into appropriate Performance Levels.

Board-Approved Cut Scores and 2013 Impact Data

The final State Board of Education approved cut scores and the percentage of Spring 2013 students in each Performance Level are shown below. These values in the scale score metric will be used for grades 4 and will not change from year to year.

Grade 4		
Performance Level	Scale Score	Percent in Category
Below	0-39	32.2%
Meets	40-56	52.3%
Exceeds	57-70	15.5%

Panelist Recruitment

The NDE recruited panelists for the Standard Setting process:

- In January of 2013, Dr. Valorie Foy communicated with District Assessment Contacts, informing them of the plan for establishing NeSA-Writing cut scores and the need for Nebraska educators to participate in the process.
- Information regarding the Standard Setting process was communicated to Nebraska districts in Nebraska's *Standards, Assessment, and Accountability Updates*.
- The NDE sought nominations for participation in the Standard Setting process.
- The NDE Statewide Assessment Office members reviewed the nominations and selected participants. Three criteria were considered:
 1. Educational role;
 2. Geographic location; and
 3. Knowledge and experience with the NeSA-Writing.
- Applicants were notified by the NDE of their selection status.

Panelist Survey

A total of 15 panelists participated in the *Body of Work* event. Table 5-1 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, and all were female.

Table 5-1 Standard Setting Panelist Summary

Demographic		Writing
Grade Group - teacher reported	4	15
Gender	Male	0
	Female	15
Ethnicity	White/non-Hispanic	14
	African American	1
Role	Other	1
	Teacher	10
	Educator	4
Region	Rural	5
	Urban	7
	Suburban	3
Experience	0 - 5 years	0
	6 - 10 years	2
	11 - 15 years	4
	16 - 20 years	1
	21 - 25 years	1
	26 - 30 years	7
	31 - 35 years	0
	> 36 years	0

Roles and Responsibilities

A successful Standard Setting requires the concerted and coordinated efforts of many people including staff from the NDE, DRC, and most importantly, the panelists. Each group has its unique and critical roles and responsibilities:

Panelists—brought their individual educational experience and expertise about Nebraska students, writing instruction, and the Nebraska curriculum. Their knowledge of writing instruction and curriculum in Nebraska and their familiarity with Nebraska students forms the foundation for the validity of the performance standards.

Nebraska Department of Education—convened the meeting and introduced the NeSA-Writing program and the importance of Standard Setting. The NDE staff monitored the progress of each panel and fielded questions on the assessment, test content, and on any policy concerns.

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

Psychometric Lead—conducted the training session and monitored progress and results throughout.

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

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

Room Facilitators—reviewed procedures for the panelists, kept the process moving on schedule, explained results, and facilitated the sessions.

Statistical Analyst—entered the panelists' ratings and performed the necessary statistical analyses.

Materials Preparation

Workshop materials were prepared by DRC. The materials available to panelists during the workshop included:

- Training materials
- Performance Level Descriptors
- Focus and Pinpointing papers
- Participant rating forms

Training materials included grade 4 writing topic responses at varying score points and related materials that were otherwise identical to the materials to be used in the actual process.

Writing Performance Level Descriptors were originally developed by the NDE with assistance from educators. A complete statement of the Performance Level Descriptors is included in Appendixes D, E, and F.

Panelists' Evaluation Results

The final step of the Standard Setting process was asking the panelists 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 G and a summary of the results as averages is shown in Table 5-2. All questions were on a four point scale unless indicated. A one represented a disagreeing statement and a four was an agreeing statement. As observed, panelists were *Confident* to *Very Confident* in the process and outcomes.

Table 5-2 Standard Setting Panelist Evaluation Results

	Grade	4
	Count	15
Training	Clarity	3.60
	Time allotted	3.20
	Exercise	3.53
PLD's	Adeq info	3.53
	Adeq time	3.57
	Capture	3.47
	Communication	3.47
Materials	PLD	3.87
	Essays	3.93
	Summary	3.67
	Impact data	3.60
Amount of time*	Focus	2.27
	Pinpointing	2.07
Roles	PS Lead	3.50
	Rm Facilitator	3.50
	Other	3.50
Confidence	Below/Meets	3.00
	Meets/Exceeds	3.43
Process	Confident	3.14

*Amount of time was on a 3 point scale where 2 was *About Right*.

REPORTING

Determining the Scale Score

The TAC felt that 200 points overstated the precision of the writing scores, because of the dominance of a few patterns. These considerations led to a choice of scale other than the 0-200 scale used by reading, math, and science. A 70-point scale was suggested, somewhat arbitrarily, as being less than 200 and different than either 50, which might be confused with a raw score, or 100, which might be confused with percent correct. Having settled on the choice of metric for the reporting scale, there is still a question of whether the weighted composite score is to be transformed linearly or logistically into the scale score. It is generally held that the logistic (Rasch) metric, when it can be used, has better measurement properties than any version of raw scores. Several Rasch analyses (multi-faceted, rating scale, weighted, unweighted) support its use with these data.

The Composite to Scale Score tables can be seen in Appendixes H and I.

Composite Scores

A composite total score is calculated from the domain scores of each reader using the weights as shown below for the four domains respectively and summing the domain scores. The composite scores will be translated into scale scores which range from 0 to 70.

The composite score for 2012 is computed by combining the domain scores as:

$$CS = 1.4D_1 + 1.0D_2 + 0.8D_3 + 0.8D_4$$

For example an 8th grade student could have received the following domain scores by reader:

	Domain 1	Domain 2	Domain 3	Domain 4	Weighted score
Reader 1	3 (4.2)	3 (3)	2 (1.6)	3 (2.4)	11.2
Reader 2	3 (4.2)	2 (2)	3 (2.4)	3 (2.4)	11.0

*Note: Weighted calculations are in parentheses.

Total composite score for this student is 22.2 which corresponds to a scale score of 40. This falls in the Performance Level *Meets the Standards*.

GRADE 4, 8, AND 11 REPORTS

DRC reported student results on the NeSA-Writing for grades 4, 8, and 11. Reports were included on the Individual Student Reports (ISRs) with NeSA- Reading, Mathematics, and Science and printed and shipped to districts/schools. Additionally, districts and schools were able to access online reports using DRC's eDIRECT system.

REFERENCES

Kingston, N. M., Kahl, S. R., Sweeney, K. P., & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting Performance Standards: Concepts, methods and perspectives* (pp. 218-248). Mahwah, NJ: Erlbaum.

Appendix A: Nebraska Department of Education Scoring Guide for Narrative Writing – Analytic – GRADE 4

Nebraska Department of Education Scoring Guide for Narrative Writing – Analytic – Grade 4				
	1	2	3	4
IDEAS / CONTENT 35%	<ul style="list-style-type: none"> The writer creates little understanding of events of the story. Content has many digressions from the topic. Supporting details are lacking. Storyline is often repetitious, disconnected, or random. 	<ul style="list-style-type: none"> The writer creates a limited understanding of events of the story. Content has some digressions from the topic. Limited or unrelated details are included. Storyline is occasionally vague. 	<ul style="list-style-type: none"> The writer creates a general understanding of events of the story. Content is generally focused on the topic. Adequate, related details are included. Storyline is generally logical and easy to follow. 	<ul style="list-style-type: none"> The writer creates a clear understanding of events of the story. Content is well-focused on the topic. Numerous, relevant details are included. Storyline is logical and easy to follow throughout.
ORGANIZATION 25%	<ul style="list-style-type: none"> Structural development of a beginning, middle, or end is lacking. Pacing is awkward. Transitions are missing or connections are unclear. Paragraphing is ineffective or missing. 	<ul style="list-style-type: none"> Structural development of a beginning, middle, or end is limited. Pacing is somewhat inconsistent. Transitions are repetitious or weak. Paragraphing is irregular. 	<ul style="list-style-type: none"> Structural development of a beginning, middle, and end is functional. Pacing is generally controlled. Transitions are functional. Paragraphing is generally successful. 	<ul style="list-style-type: none"> Structural development of a beginning, middle, and end is effective. Pacing is well-controlled. Transitions effectively show how ideas connect. Paragraphing is sound.
VOICE / WORD CHOICE 20%	<ul style="list-style-type: none"> Wording is lifeless and mechanical, conveying little sense of the writer. Voice is inappropriate for the purpose and audience. Language is neither specific, precise, nor varied. 	<ul style="list-style-type: none"> Wording is occasionally expressive, conveying a limited sense of the writer. Voice is sometimes inappropriate for the purpose and audience. Language is occasionally specific, precise, and varied. 	<ul style="list-style-type: none"> Wording is generally expressive, conveying a sense of the writer. Voice is generally appropriate for the purpose and audience. Language is generally specific, precise, and varied. 	<ul style="list-style-type: none"> Wording is expressive and engaging, conveying a strong sense of the writer. Voice is well-suited for the purpose and audience. Language is specific, precise, and varied throughout.
SENTENCE FLUENCY / CONVENTIONS 20%	<ul style="list-style-type: none"> Sentences seldom vary in length or structure. Phrasing sounds awkward and unnatural. Fragments or run-ons confuse the reader. Grammar, usage, punctuation, and spelling errors throughout distract the reader. 	<ul style="list-style-type: none"> Sentences occasionally vary in length or structure. Phrasing occasionally sounds unnatural. Fragments or run-ons sometimes confuse the reader. Grammar, usage, punctuation, and spelling errors may distract the reader. 	<ul style="list-style-type: none"> Sentences generally vary in length or structure. Phrasing generally sounds natural. Fragments and run-ons, if present, do not confuse the reader. Grammar, usage, punctuation, and spelling are usually correct and errors do not distract the reader. 	<ul style="list-style-type: none"> Sentences vary in length and structure throughout. Phrasing consistently sounds natural and conveys meaning. Fragments and run-ons, if present, are intended for stylistic effect. Grammar, usage, punctuation, and spelling are consistently correct and may be manipulated for stylistic effect.

Appendix B: Nebraska Department of Education Scoring Guide for Descriptive Writing – Analytic – GRADE 8

Nebraska Department of Education Scoring Guide for Descriptive Writing – Analytic - GRADE 8				
	1	2	3	4
IDEAS / CONTENT 35%	<ul style="list-style-type: none"> The picture of what is being described is unclear. Content has many digressions from the topic. Sensory details are lacking. 	<ul style="list-style-type: none"> The picture of what is being described is limited. Content has some digressions from the topic. Sensory details are limited or unrelated. 	<ul style="list-style-type: none"> The picture of what is being described is clear. Content is generally focused on the topic. Sensory details are adequate and related. 	<ul style="list-style-type: none"> The picture of what is being described is clear and vivid. Content is well-focused on the topic. Sensory details are numerous and relevant.
ORGANIZATION 25%	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is lacking. Pacing is awkward. Transitions are missing or connections are unclear. Paragraphing is ineffective or missing. 	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is limited. Pacing is somewhat inconsistent. Transitions are repetitious or weak. Paragraphing is irregular. 	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is functional. Pacing is generally controlled. Transitions are functional. Paragraphing is generally successful. 	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is effective. Pacing is well-controlled. Transitions effectively show how ideas connect. Paragraphing is sound.
VOICE / WORD CHOICE 20%	<ul style="list-style-type: none"> Wording is inexpressive and lifeless, conveying little sense of the writer. Voice inappropriate for the purpose and audience. Language is neither specific, precise, nor varied. Few, if any, vivid words or phrases are used. 	<ul style="list-style-type: none"> Wording is occasionally expressive, conveying a limited sense of the writer. Voice is sometimes inappropriate for the purpose and audience. Language is occasionally specific, precise, and varied. Some vivid words and phrases are used. 	<ul style="list-style-type: none"> Wording is generally expressive, conveying a sense of the writer. Voice is generally appropriate for the purpose and audience. Language is generally specific, precise, and varied. Adequate vivid words and phrases are used. 	<ul style="list-style-type: none"> Wording is expressive and engaging, conveying a strong sense of the writer throughout. Voice is well-suited for the purpose and audience throughout. Language is specific, precise, and varied throughout. Numerous vivid words and phrases used effectively.
SENTENCE FLUENCY / CONVENTIONS 20%	<ul style="list-style-type: none"> Sentences seldom vary in length or structure. Phrasing sounds awkward and unnatural. Fragments or run-ons confuse the reader. Grammar, usage, punctuation, and spelling errors throughout distract the reader. 	<ul style="list-style-type: none"> Sentences occasionally vary in length or structure. Phrasing occasionally sounds unnatural. Fragments or run-ons sometimes confuse the reader. Grammar, usage, punctuation, and spelling errors may distract the reader. 	<ul style="list-style-type: none"> Sentences generally vary in length or structure. Phrasing generally sounds natural. Fragments and run-ons, if present, do not confuse the reader. Grammar, usage, punctuation, and spelling are usually correct and errors do not distract the reader. 	<ul style="list-style-type: none"> Sentences vary in length and structure throughout. Phrasing consistently sounds natural and conveys meaning. Fragments and run-ons, if present, are intended for stylistic effect. Grammar, usage, punctuation, and spelling are consistently correct and may be manipulated for stylistic effect.

Appendix C: Nebraska Department of Education Scoring Guide for Persuasive Writing – Analytic – GRADE 11

Nebraska Department of Education Scoring Guide for Persuasive Writing – Analytic – GRADE 11				
	1	2	3	4
IDEAS / CONTENT 35%	<ul style="list-style-type: none"> Writer conveys little opinion or position about the topic. Content has many digressions from the topic. Reasoning is unclear. Supporting examples or reasons are lacking. 	<ul style="list-style-type: none"> Writer conveys a limited opinion or position about the topic. Content has some digressions from the topic. Reasoning is somewhat logical and convincing. Supporting examples or reasons are limited. 	<ul style="list-style-type: none"> Writer conveys a general opinion or position about the topic. Content is generally focused on the topic. Reasoning is usually logical and convincing. Supporting examples or reasons are adequate and relevant. 	<ul style="list-style-type: none"> Writer conveys a clear opinion or position about the topic. Content is well-focused on the topic. Reasoning is logical and compelling. Supporting examples or reasons are numerous and relevant.
ORGANIZATION 25%	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is lacking. Pacing is awkward. Transitions are missing or connections are unclear. Paragraphing is ineffective or missing. 	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is limited. Pacing is somewhat inconsistent. Transitions are repetitious or weak. Paragraphing is irregular. 	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is functional. Pacing is generally controlled. Transitions are functional. Paragraphing is generally successful. 	<ul style="list-style-type: none"> Structural development of an introduction, body, and conclusion is effective. Pacing is well- controlled. Transitions effectively show how ideas connect. Paragraphing is sound.
VOICE / WORD CHOICE 20%	<ul style="list-style-type: none"> Writer demonstrates little commitment to the topic. Voice is inappropriate for the purpose and audience. Language is neither specific, precise, varied, nor engaging. Writer fails to anticipate the reader's questions. 	<ul style="list-style-type: none"> Writer demonstrates a limited commitment to the topic. Voice is sometimes inappropriate for the purpose and audience. Language is occasionally specific, precise, varied, and engaging. Writer anticipates few of the reader's questions. 	<ul style="list-style-type: none"> Writer demonstrates a general commitment to the topic. Voice is generally appropriate for the purpose and audience. Language is generally specific, precise, varied, and engaging. Writer generally anticipates the reader's questions. 	<ul style="list-style-type: none"> Writer demonstrates a strong commitment to the topic. Voice is well-suited for the purpose and audience. Language is specific, precise, varied, and engaging throughout. Writer consistently anticipates reader's questions.
SENTENCE FLUENCY / CONVENTIONS 20%	<ul style="list-style-type: none"> Sentences seldom vary in length or structure. Phrasing sounds awkward and unnatural. Fragment or run-ons confuse the reader. Grammar, usage, punctuation, and spelling errors throughout distract the reader. 	<ul style="list-style-type: none"> Sentences occasionally vary in length or structure. Phrasing occasionally sounds unnatural. Fragments or run-ons sometimes confuse the reader. Grammar, usage, punctuation, and spelling errors may distract the reader. 	<ul style="list-style-type: none"> Sentences generally vary in length or structure. Phrasing generally sounds natural. Fragments and run-ons, if present, do not confuse the reader. Grammar, usage, punctuation, and spelling are usually correct and errors do not distract the reader. 	<ul style="list-style-type: none"> Sentences vary in length and structure throughout. Phrasing consistently sounds natural and conveys meaning. Fragments and run-ons, if present, are intended for stylistic effect. Grammar, usage, punctuation, and spelling are consistently correct and may be manipulated for stylistic effect.

Appendix D: Performance Level Descriptors Grade 4

Nebraska State Accountability-Writing (NeSA-W) Performance Level Descriptors Grade 4

<u>Below the Standards</u>	<u>Meets the Standards</u>	<u>Exceeds the Standards</u>
<p>Overall the student's writing reflects an unsatisfactory performance of the standards and an insufficient understanding of the traits of writing. The student's writing is still under development. Extensive revision and/or editing is necessary.</p> <p>The student's writing is below the standards if the . . .</p> <ul style="list-style-type: none"> o Writer creates a limited or no understanding of events in the story. o Content has some digressions from the topic. o Supporting details are limited, unrelated, or lacking. o Storyline is vague, repetitious, disconnected, or random. o Structural development of a beginning, middle, or end is limited or lacking. o Pacing is inconsistent or awkward. o Transitions are repetitious, weak, unclear, or missing. o Paragraphing is irregular, ineffective, or missing. o Wording is inexpressive and lifeless, conveying a limited sense of the writer. o Voice is sometimes inappropriate for the purpose and audience. o Language is seldom specific, precise or varied. o Sentences seldom vary in length or structure. o Phrasing sounds awkward and unnatural. o Writing has fragments or run-ons that confuse the reader. o Grammar, usage, punctuation, and spelling errors distract the reader. 	<p>Overall the student's writing reflects a satisfactory performance of the standards and a sufficient understanding of the traits of writing. The student's writing demonstrates more strengths than weaknesses. Some revision and/or editing is necessary.</p> <p>The student's writing meets the standards if the . . .</p> <ul style="list-style-type: none"> o Writer creates a general understanding of events in the story. o Content is generally focused on the topic. o Details are adequate and related. o Storyline is generally logical and easy to follow. o Structural development of a beginning, middle, and end is functional. o Pacing is generally controlled. o Transitions are functional. o Paragraphing is generally successful. o Wording is generally expressive, conveying a sense of the writer. o Voice is generally appropriate for the purpose and audience. o Language is generally specific, precise, and varied. o Sentences generally vary in length or structure. o Phrasing generally sounds natural. o Fragments and run-ons do not generally confuse the reader. o Grammar, usage, punctuation, and spelling are usually correct and rarely distract the reader. 	<p>Overall the student's writing reflects an advanced performance of the standards and a thorough understanding of the traits of writing. The student's writing demonstrates numerous strengths. Only minor revision and/or editing is necessary.</p> <p>The student's writing exceeds the standards if the . . .</p> <ul style="list-style-type: none"> o Writer creates a clear understanding of events in the story. o Content is well-focused on the topic. o Details are numerous and relevant. o Storyline is logical and easy to follow throughout. o Structural development of a beginning, middle, and end is effective. o Pacing is well-controlled. o Transitions effectively show how ideas connect. o Paragraphing is sound. o Wording is expressive and engaging, conveying a strong sense of the writer throughout. o Voice is well-suited for the purpose and audience throughout. o Language is specific, precise, and varied throughout. o Sentences vary in length and structure throughout. o Phrasing consistently sounds natural and conveys meaning. o Fragments and run-ons, if present, are intended for stylistic effect. o Grammar, usage, punctuation, and spelling are consistently correct and may be manipulated for stylistic effect.

Appendix E: Performance Level Descriptors Grade 8

Nebraska State Accountability-Writing (NeSA-W) Performance Level Descriptors

Grade 8

<u>Below the Standards</u>	<u>Meets the Standards</u>	<u>Exceeds the Standards</u>
<p>Overall the student's writing reflects an unsatisfactory performance of the standards and an insufficient understanding of the traits of writing. The student's writing is still under development. Extensive revision and/or editing is necessary.</p> <p>The student's writing is below the standards if the . . .</p> <ul style="list-style-type: none"> ○ Picture of what is being described is limited or unclear. ○ Content has some digressions from the topic. ○ Sensory details are limited, unrelated, or lacking. ○ Structural development of an introduction, body, and conclusion is limited or lacking. ○ Pacing is inconsistent or awkward. ○ Transitions are repetitious, weak, unclear, or missing. ○ Paragraphing is irregular, ineffective, or missing. ○ Wording is inexpressive and lifeless, conveying a limited sense of the writer. ○ Voice is sometimes inappropriate for the purpose and audience. ○ Language is seldom specific, precise or varied. ○ Writing lacks vivid words and phrases ○ Sentences seldom vary in length or structure. ○ Phrasing sounds awkward and unnatural. ○ Writing has fragments or run-ons that confuse the reader. ○ Grammar, usage, punctuation, and spelling errors distract the reader. 	<p>Overall the student's writing reflects a satisfactory performance of the standards and a sufficient understanding of the traits of writing. The student's writing demonstrates more strengths than weaknesses. Some revision and/or editing is necessary.</p> <p>The student's writing meets the standards if the . . .</p> <ul style="list-style-type: none"> ○ Picture of what is being described is clear. ○ Content is generally focused on the topic. ○ Sensory details are adequate and related. ○ Structural development of an introduction, body, and conclusion is functional. ○ Pacing is generally controlled. ○ Transitions are functional. ○ Paragraphing is generally successful. ○ Wording is generally expressive, conveying a sense of the writer. ○ Voice is generally appropriate for the purpose and audience. ○ Language is generally specific, precise, and varied. ○ Writing has adequate vivid words and phrases. ○ Sentences generally vary in length or structure. ○ Phrasing generally sounds natural. ○ Fragments and run-ons do not generally confuse the reader. ○ Grammar, usage, punctuation, and spelling are usually correct and rarely distract the reader. 	<p>Overall the student's writing reflects an advanced performance of the standards and a thorough understanding of the traits of writing. The student's writing demonstrates numerous strengths. Only minor revision and/or editing is necessary.</p> <p>The student's writing exceeds the standards if the . . .</p> <ul style="list-style-type: none"> ○ Picture of what is being described is clear and vivid. ○ Content is well-focused on the topic. ○ Sensory details are numerous and relevant. ○ Structural development of an introduction, body, and conclusion is effective. ○ Pacing is well-controlled. ○ Transitions effectively show how ideas connect. ○ Paragraphing is sound. ○ Wording is expressive and engaging, conveying a strong sense of the writer throughout. ○ Voice is well-suited for the purpose and audience throughout. ○ Language is specific, precise, and varied throughout. ○ Numerous vivid words and phrases are used effectively. ○ Sentences vary in length and structure throughout. ○ Phrasing consistently sounds natural and conveys meaning. ○ Fragments and run-ons, if present, are intended for stylistic effect. ○ Grammar, usage, punctuation, and spelling are consistently correct and may be manipulated for stylistic effect.

Appendix F: Performance Level Descriptors Grade 11

Nebraska State Accountability-Writing (NeSA-W) Performance Level Descriptors

Grade 11

Below the Standards

Overall the student's writing reflects an unsatisfactory performance of the standards and an insufficient understanding of the traits of writing. The student's writing is still under development. Extensive revision and/or editing is necessary.

The student's writing is below the standards if the . . .

- Writer conveys limited or no opinion or position about the topic.
- Content has some digressions from the topic.
- Reasoning is limited or unclear.
- Supporting examples or reasons are limited or lacking.
- Structural development of an introduction, body, and conclusion is limited or lacking.
- Pacing is inconsistent or awkward.
- Transitions are repetitious, weak, unclear, or missing.
- Paragraphing is irregular, ineffective, or missing.
- Writer demonstrates limited or no commitment to the topic.
- Voice is sometimes inappropriate for the purpose and audience.
- Language is seldom specific, precise, or varied.
- Writer often fails to anticipate the reader's questions.
- Sentences seldom vary in length or structure.
- Phrasing sounds awkward and unnatural.
- Writing includes fragments or run-ons that confuse the reader.
- Grammar, usage, punctuation, and spelling errors distract the reader.

Meets the Standards

Overall the student's writing reflects a satisfactory performance of the standards and a sufficient understanding of the traits of writing. The student's writing demonstrates more strengths than weaknesses. Some revision and/or editing is necessary.

The student's writing meets the standards if the . . .

- Writer conveys a general opinion or position about the topic.
- Content is generally focused on the topic.
- Reasoning is usually logical and convincing.
- Supporting examples or reasons are adequate and relevant.
- Structural development of an introduction, body, and conclusion is functional.
- Pacing is generally controlled.
- Transitions are functional.
- Paragraphing is generally successful.
- Writer demonstrates a general commitment to the topic.
- Voice is generally appropriate for the purpose and audience.
- Language is generally specific, precise, varied, and engaging.
- Writer generally anticipates the reader's questions.
- Sentences generally vary in length or structure.
- Phrasing generally sounds natural.
- Fragments and run-ons, if present, generally do not confuse the reader.
- Grammar, usage, punctuation, and spelling are usually correct and errors rarely distract the reader.

Exceeds the Standards

Overall the student's writing reflects an advanced performance of the standards and a thorough understanding of the traits of writing. The student's writing demonstrates numerous strengths. Only minor revision and/or editing is necessary.

The student's writing exceeds the standards if the . . .

- Writer conveys a clear opinion or position about the topic.
- Content is well-focused on the topic.
- Reasoning is logical and compelling.
- Supporting examples or reasons are numerous and relevant.
- Structural development of an introduction, body, and conclusion is effective.
- Pacing is well-controlled.
- Transitions effectively show how ideas connect.
- Paragraphing is sound.
- Writer demonstrates a strong commitment to the topic.
- Voice is well-suited for the purpose and audience.
- Language is specific, precise, varied, and engaging throughout.
- Writer consistently anticipates reader's questions.
- Sentences vary in length and structure throughout.
- Phrasing consistently sounds natural and conveys meaning.
- Fragments and run-ons, if present, are intended for stylistic effect.
- Grammar, usage, punctuation, and spelling are consistently correct and may be manipulated for stylistic effect.

Appendix G: Standard Setting Panelist Evaluation Form

NEBRASKA STATE ACCOUNTABILITY-WRITING (NESA-W)
STANDARD SETTING MEETING
MARCH 21, 2013
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. AND ALSO NOTE, IN ORDER FOR YOUR ANSWERS TO BE INCLUDED PLEASE CLEARLY STATE YOUR RESPONSE.

1. Grade Level:

4

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 Performance Level Descriptors (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 Below the Standards, Meets the Standards, and Exceeds the Standards.				

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

Materials	Not Useful	Somewhat Useful	Useful	Very Useful
PLD's				
Essays				
Panelist Summary				
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
Focus Round			
Pinpointing Round			

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

Role	Not Satisfied	Somewhat 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
4	Below/Meets				
	Meets/Exceeds				

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

Not Confident Somewhat Confident Confident Very Confident

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

THANK YOU FOR PARTICIPATING IN THE STANDARD SETTING MEETING.

Appendix H: Composite to Scale Score Tables Grade 4

Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score
4	8.0	1	4	12.0	19	4	16.0	29
4	8.1	7	4	12.1	19	4	16.1	31
4	8.2	10	4	12.2	19	4	16.2	32
4	8.3	11	4	12.3	19	4	16.3	32
4	8.4	12	4	12.4	20	4	16.4	33
4	8.5	13	4	12.5	20	4	16.5	33
4	8.6	13	4	12.6	20	4	16.6	34
4	8.7	13	4	12.7	20	4	16.7	34
4	8.8	14	4	12.8	20	4	16.8	35
4	8.9	14	4	12.9	20	4	16.9	35
4	9.0	14	4	13.0	20	4	17.0	35
4	9.1	15	4	13.1	20	4	17.1	35
4	9.2	15	4	13.2	21	4	17.2	36
4	9.3	15	4	13.3	21	4	17.3	36
4	9.4	15	4	13.4	21	4	17.4	36
4	9.5	16	4	13.5	21	4	17.5	36
4	9.6	16	4	13.6	21	4	17.6	36
4	9.7	16	4	13.7	21	4	17.7	37
4	9.8	16	4	13.8	21	4	17.8	37
4	9.9	16	4	13.9	22	4	17.9	37
4	10.0	16	4	14.0	22	4	18.0	37
4	10.1	17	4	14.1	22	4	18.1	37
4	10.2	17	4	14.2	22	4	18.2	37
4	10.3	17	4	14.3	22	4	18.3	38
4	10.4	17	4	14.4	22	4	18.4	38
4	10.5	17	4	14.5	23	4	18.5	38
4	10.6	17	4	14.6	23	4	18.6	38
4	10.7	17	4	14.7	23	4	18.7	38
4	10.8	18	4	14.8	23	4	18.8	38
4	10.9	18	4	14.9	23	4	18.9	38
4	11.0	18	4	15.0	24	4	19.0	39
4	11.1	18	4	15.1	24	4	19.1	39
4	11.2	18	4	15.2	24	4	19.2	39
4	11.3	18	4	15.3	25	4	19.3	39
4	11.4	18	4	15.4	25	4	19.4	39
4	11.5	18	4	15.5	25	4	19.5	39
4	11.6	19	4	15.6	26	4	19.6	39
4	11.7	19	4	15.7	27	4	19.7	39
4	11.8	19	4	15.8	27	4	19.8	40
4	11.9	19	4	15.9	28	4	19.9	40

Nebraska State Accountability 2013 Writing Technical Report

Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score
4	20.0	40	4	24.0	50	4	28.0	61
4	20.1	40	4	24.1	51	4	28.1	61
4	20.2	40	4	24.2	52	4	28.2	61
4	20.3	40	4	24.3	53	4	28.3	61
4	20.4	40	4	24.4	54	4	28.4	61
4	20.5	40	4	24.5	54	4	28.5	61
4	20.6	40	4	24.6	55	4	28.6	61
4	20.7	41	4	24.7	55	4	28.7	61
4	20.8	41	4	24.8	55	4	28.8	62
4	20.9	41	4	24.9	56	4	28.9	62
4	21.0	41	4	25.0	56	4	29.0	62
4	21.1	41	4	25.1	56	4	29.1	62
4	21.2	41	4	25.2	57	4	29.2	62
4	21.3	41	4	25.3	57	4	29.3	62
4	21.4	41	4	25.4	57	4	29.4	62
4	21.5	42	4	25.5	57	4	29.5	63
4	21.6	42	4	25.6	57	4	29.6	63
4	21.7	42	4	25.7	58	4	29.7	63
4	21.8	42	4	25.8	58	4	29.8	63
4	21.9	42	4	25.9	58	4	29.9	63
4	22.0	42	4	26.0	58	4	30.0	63
4	22.1	43	4	26.1	58	4	30.1	63
4	22.2	43	4	26.2	58	4	30.2	64
4	22.3	43	4	26.3	58	4	30.3	64
4	22.4	43	4	26.4	59	4	30.4	64
4	22.5	43	4	26.5	59	4	30.5	64
4	22.6	43	4	26.6	59	4	30.6	64
4	22.7	44	4	26.7	59	4	30.7	65
4	22.8	44	4	26.8	59	4	30.8	65
4	22.9	44	4	26.9	59	4	30.9	65
4	23.0	44	4	27.0	59	4	31.0	65
4	23.1	45	4	27.1	60	4	31.1	66
4	23.2	45	4	27.2	60	4	31.2	66
4	23.3	45	4	27.3	60	4	31.3	66
4	23.4	46	4	27.4	60	4	31.4	67
4	23.5	46	4	27.5	60	4	31.5	67
4	23.6	47	4	27.6	60	4	31.6	68
4	23.7	47	4	27.7	60	4	31.7	68
4	23.8	48	4	27.8	60	4	31.8	69
4	23.9	49	4	27.9	61	4	31.9	70
						4	32.0	70

Appendix I: Composite to Scale Score Tables Grade 8

Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score
8	8.0	1	8	12.0	16	8	16.0	27
8	8.1	7	8	12.1	16	8	16.1	28
8	8.2	8	8	12.2	16	8	16.2	29
8	8.3	9	8	12.3	17	8	16.3	30
8	8.4	10	8	12.4	17	8	16.4	30
8	8.5	10	8	12.5	17	8	16.5	31
8	8.6	11	8	12.6	17	8	16.6	31
8	8.7	11	8	12.7	17	8	16.7	32
8	8.8	11	8	12.8	17	8	16.8	32
8	8.9	12	8	12.9	17	8	16.9	32
8	9.0	12	8	13.0	17	8	17.0	33
8	9.1	12	8	13.1	17	8	17.1	33
8	9.2	12	8	13.2	18	8	17.2	33
8	9.3	13	8	13.3	18	8	17.3	33
8	9.4	13	8	13.4	18	8	17.4	34
8	9.5	13	8	13.5	18	8	17.5	34
8	9.6	13	8	13.6	18	8	17.6	34
8	9.7	13	8	13.7	18	8	17.7	34
8	9.8	14	8	13.8	18	8	17.8	34
8	9.9	14	8	13.9	18	8	17.9	34
8	10.0	14	8	14.0	19	8	18.0	35
8	10.1	14	8	14.1	19	8	18.1	35
8	10.2	14	8	14.2	19	8	18.2	35
8	10.3	14	8	14.3	19	8	18.3	35
8	10.4	14	8	14.4	19	8	18.4	35
8	10.5	15	8	14.5	19	8	18.5	35
8	10.6	15	8	14.6	20	8	18.6	36
8	10.7	15	8	14.7	20	8	18.7	36
8	10.8	15	8	14.8	20	8	18.8	36
8	10.9	15	8	14.9	20	8	18.9	36
8	11.0	15	8	15.0	20	8	19.0	36
8	11.1	15	8	15.1	21	8	19.1	36
8	11.2	15	8	15.2	21	8	19.2	36
8	11.3	15	8	15.3	21	8	19.3	36
8	11.4	16	8	15.4	22	8	19.4	37
8	11.5	16	8	15.5	22	8	19.5	37
8	11.6	16	8	15.6	23	8	19.6	37
8	11.7	16	8	15.7	23	8	19.7	37
8	11.8	16	8	15.8	24	8	19.8	37
8	11.9	16	8	15.9	25	8	19.9	37

Nebraska State Accountability 2013 Writing Technical Report

Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score
8	20.0	37	8	24.0	48	8	28.0	61
8	20.1	37	8	24.1	50	8	28.1	61
8	20.2	37	8	24.2	51	8	28.2	61
8	20.3	38	8	24.3	52	8	28.3	61
8	20.4	38	8	24.4	53	8	28.4	61
8	20.5	38	8	24.5	53	8	28.5	61
8	20.6	38	8	24.6	54	8	28.6	62
8	20.7	38	8	24.7	54	8	28.7	62
8	20.8	38	8	24.8	55	8	28.8	62
8	20.9	38	8	24.9	55	8	28.9	62
8	21.0	38	8	25.0	55	8	29.0	62
8	21.1	39	8	25.1	55	8	29.1	62
8	21.2	39	8	25.2	56	8	29.2	62
8	21.3	39	8	25.3	56	8	29.3	63
8	21.4	39	8	25.4	56	8	29.4	63
8	21.5	39	8	25.5	56	8	29.5	63
8	21.6	39	8	25.6	57	8	29.6	63
8	21.7	39	8	25.7	57	8	29.7	63
8	21.8	40	8	25.8	57	8	29.8	63
8	21.9	40	8	25.9	57	8	29.9	64
8	22.0	40	8	26.0	57	8	30.0	64
8	22.1	40	8	26.1	58	8	30.1	64
8	22.2	40	8	26.2	58	8	30.2	64
8	22.3	40	8	26.3	58	8	30.3	64
8	22.4	40	8	26.4	58	8	30.4	65
8	22.5	41	8	26.5	58	8	30.5	65
8	22.6	41	8	26.6	58	8	30.6	65
8	22.7	41	8	26.7	59	8	30.7	65
8	22.8	41	8	26.8	59	8	30.8	66
8	22.9	42	8	26.9	59	8	30.9	66
8	23.0	42	8	27.0	59	8	31.0	66
8	23.1	42	8	27.1	59	8	31.1	66
8	23.2	42	8	27.2	59	8	31.2	67
8	23.3	43	8	27.3	60	8	31.3	67
8	23.4	43	8	27.4	60	8	31.4	67
8	23.5	43	8	27.5	60	8	31.5	68
8	23.6	44	8	27.6	60	8	31.6	68
8	23.7	45	8	27.7	60	8	31.7	69
8	23.8	45	8	27.8	60	8	31.8	70
8	23.9	47	8	27.9	60	8	31.9	70
						8	32.0	70

Appendix J: Composite to Scale Score Tables Grade 11

Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score
11	8.0	1	11	12.0	17	11	16.0	25
11	8.1	5	11	12.1	17	11	16.1	26
11	8.2	7	11	12.2	17	11	16.2	26
11	8.3	8	11	12.3	17	11	16.3	27
11	8.4	9	11	12.4	17	11	16.4	27
11	8.5	9	11	12.5	17	11	16.5	28
11	8.6	10	11	12.6	17	11	16.6	28
11	8.7	10	11	12.7	18	11	16.7	28
11	8.8	11	11	12.8	18	11	16.8	29
11	8.9	11	11	12.9	18	11	16.9	29
11	9.0	11	11	13.0	18	11	17.0	29
11	9.1	11	11	13.1	18	11	17.1	29
11	9.2	12	11	13.2	18	11	17.2	30
11	9.3	12	11	13.3	18	11	17.3	30
11	9.4	12	11	13.4	19	11	17.4	30
11	9.5	12	11	13.5	19	11	17.5	30
11	9.6	13	11	13.6	19	11	17.6	31
11	9.7	13	11	13.7	19	11	17.7	31
11	9.8	13	11	13.8	19	11	17.8	31
11	9.9	13	11	13.9	19	11	17.9	31
11	10.0	13	11	14.0	20	11	18.0	31
11	10.1	14	11	14.1	20	11	18.1	31
11	10.2	14	11	14.2	20	11	18.2	32
11	10.3	14	11	14.3	20	11	18.3	32
11	10.4	14	11	14.4	20	11	18.4	32
11	10.5	14	11	14.5	21	11	18.5	32
11	10.6	14	11	14.6	21	11	18.6	32
11	10.7	15	11	14.7	21	11	18.7	32
11	10.8	15	11	14.8	21	11	18.8	33
11	10.9	15	11	14.9	21	11	18.9	33
11	11.0	15	11	15.0	22	11	19.0	33
11	11.1	15	11	15.1	22	11	19.1	33
11	11.2	15	11	15.2	22	11	19.2	33
11	11.3	16	11	15.3	23	11	19.3	33
11	11.4	16	11	15.4	23	11	19.4	33
11	11.5	16	11	15.5	23	11	19.5	34
11	11.6	16	11	15.6	24	11	19.6	34
11	11.7	16	11	15.7	24	11	19.7	34
11	11.8	16	11	15.8	25	11	19.8	34
11	11.9	16	11	15.9	25	11	19.9	34

Nebraska State Accountability 2013 Writing Technical Report

Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score	Grade	Composite Score	Scale Score
11	20.0	34	11	24.0	46	11	28.0	59
11	20.1	34	11	24.1	47	11	28.1	60
11	20.2	35	11	24.2	48	11	28.2	60
11	20.3	35	11	24.3	49	11	28.3	60
11	20.4	35	11	24.4	50	11	28.4	60
11	20.5	35	11	24.5	51	11	28.5	61
11	20.6	35	11	24.6	51	11	28.6	61
11	20.7	35	11	24.7	51	11	28.7	61
11	20.8	35	11	24.8	52	11	28.8	61
11	20.9	36	11	24.9	52	11	28.9	61
11	21.0	36	11	25.0	53	11	29.0	62
11	21.1	36	11	25.1	53	11	29.1	62
11	21.2	36	11	25.2	53	11	29.2	62
11	21.3	36	11	25.3	53	11	29.3	62
11	21.4	36	11	25.4	54	11	29.4	62
11	21.5	36	11	25.5	54	11	29.5	63
11	21.6	37	11	25.6	54	11	29.6	63
11	21.7	37	11	25.7	54	11	29.7	63
11	21.8	37	11	25.8	55	11	29.8	63
11	21.9	37	11	25.9	55	11	29.9	64
11	22.0	37	11	26.0	55	11	30.0	64
11	22.1	37	11	26.1	55	11	30.1	64
11	22.2	38	11	26.2	55	11	30.2	64
11	22.3	38	11	26.3	56	11	30.3	64
11	22.4	38	11	26.4	56	11	30.4	65
11	22.5	38	11	26.5	56	11	30.5	65
11	22.6	39	11	26.6	56	11	30.6	65
11	22.7	39	11	26.7	57	11	30.7	65
11	22.8	39	11	26.8	57	11	30.8	66
11	22.9	39	11	26.9	57	11	30.9	66
11	23.0	40	11	27.0	57	11	31.0	66
11	23.1	40	11	27.1	57	11	31.1	67
11	23.2	40	11	27.2	58	11	31.2	67
11	23.3	40	11	27.3	58	11	31.3	67
11	23.4	41	11	27.4	58	11	31.4	68
11	23.5	41	11	27.5	58	11	31.5	68
11	23.6	42	11	27.6	59	11	31.6	69
11	23.7	43	11	27.7	59	11	31.7	70
11	23.8	43	11	27.8	59	11	31.8	70
11	23.9	45	11	27.9	59	11	31.9	70
						11	32.0	70

Appendix K: Formatting Information for TAC

NeSA Writing 2013 - Formatting Analysis

Nebraska NeSA Accountability TAC - Presented by Bill Auty - April 24, 2013

The following analyses are based on identification of formatting issues by raters. If both raters identified a paper as showing evidence of formatting issues caused by the online software, the paper is flagged. Note that NDE believes that the number of papers flagged may be a slight overestimate of the online formatting issue. For example, some students may have intentionally entered unusual spacings and returns, which would look to the raters as if the software caused the formatting.

DF\$Grade	FALSE	TRUE	
08	11458	9490	20948
	54.7	45.3	
11	11952	8869	20821
	57.4	42.6	
Total	23410	18359	41769

Table 1: Papers Flagged for Formatting by Grade

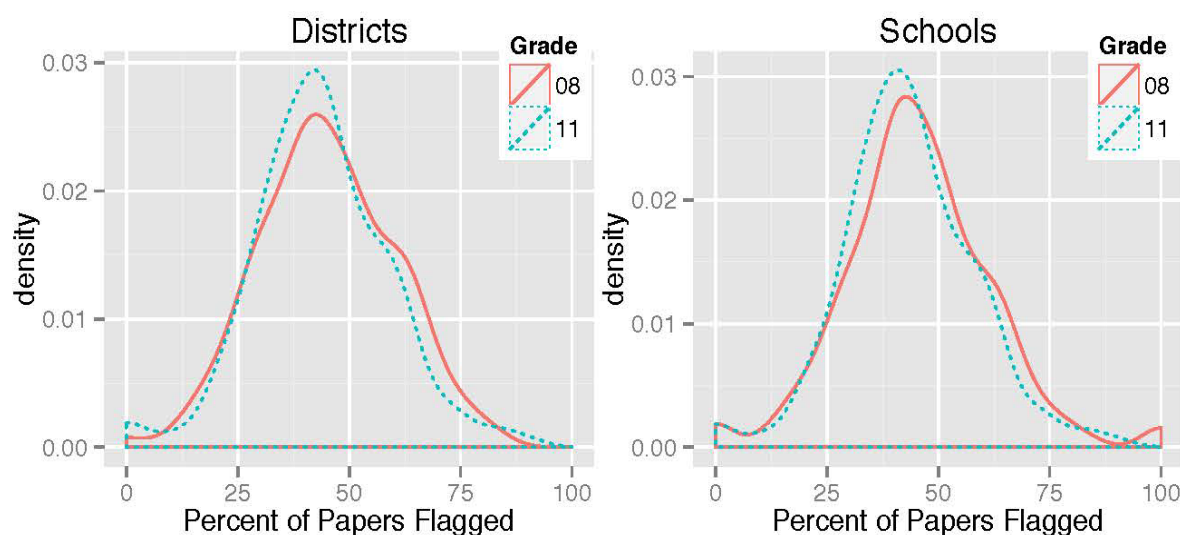


Figure 1: Distribution of Percent Flagged

These results lead us to conclude that the formatting issue was not limited to a few students or a few schools. Across the state at grade 8, 45% of the papers were flagged for formatting and 43% were flagged at grade 11. The distributions of schools and districts by percent of papers flagged for formatting also indicate that the issue was wide spread across districts and schools.

Grade 8

To test the effect of the formatting issue on writing results, we conducted the following analyses:

- A comparison of means of the domains and total scores of papers flagged and not flagged
- ANOVA for the effect of formatting flag on total writing score
- ANCOVA for the total writing score by grade 7 reading and formatting flag
- A graphic representation of those data
- A comparison of the correlations between domain scores
- A comparison of the mean scores by percent of papers flagged at the school and district level

flag	Grade	IC	ORG	VWC	SFC	Total
FALSE	08	5.58	5.50	5.64	5.52	22.24
TRUE	08	5.85	5.71	5.88	5.68	23.14

Table 2: Writing Score Means by Format Flag

The mean scores of papers flagged for formatting are all higher than of the means of papers not flagged.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
flag	1.000	4205.272	4205.272	172.849	0.000
Residuals	20946.000	509599.088	24.329		

Table 3: ANOVA: Total by Format Flag

The Analysis of Variance (ANOVA) results show a significant main effect of difference in mean total writing score between the flagged and non-flagged papers.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
flag	1.000	3356.732	3356.732	202.234	0.000
read7	1.000	130766.352	130766.352	7878.309	0.000
flag:read7	1.000	109.612	109.612	6.604	0.010
Residuals	19998.000	331932.323	16.598		

Table 4: ANCOVA: Total by Format Flag on Grade 7 Reading

The analysis of covariance (ANCOVA) was run to see if the differences in writing scores were due to the ability of students in the group with formatting issues. We don't have any concurrent NESA results for this year, so we used the scale score on NESA-Reading from last year as an indicator of student ability. The results of the analysis indicate that previous year's reading scores are highly related to this year's writing results and that the differences between flagged and non-flagged papers were significant even when grade 7 reading was taken into account. The interaction term was also significant. The graph in figure 2 shows that papers flagged for formatting had higher scores across the range of grade 7 reading scale scores with the difference diminishing at the upper end of the range.

The correlation matrices in tables 6 and 6 show high correlations between domains. The Chi Squared test for the equality of correlation matrices has a value of 1872.341 (probability ≤ 0) indicating that the correlations are different with the flagged papers having lower correlations. However, inspection of the matrices reveals that the patterns of correlations are the same, i.e., the differences are not due to a single domain.

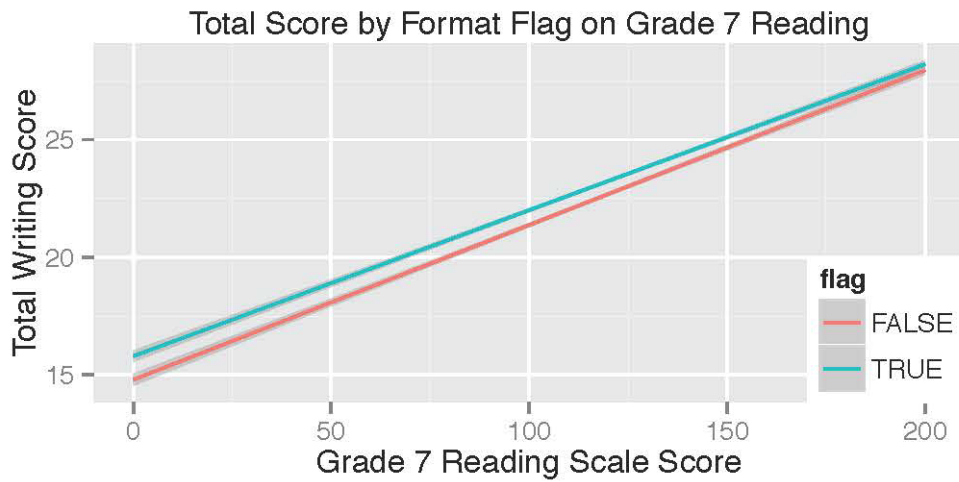


Figure 2: ANCOVA Plot

	IC	ORG	VWC	SFC
IC	1.00	0.90	0.89	0.85
ORG	0.90	1.00	0.87	0.87
VWC	0.89	0.87	1.00	0.88
SFC	0.85	0.87	0.88	1.00

Table 5: Domain Score Correlations: Papers Not Flagged

	IC	ORG	VWC	SFC
IC	1.00	0.85	0.85	0.80
ORG	0.85	1.00	0.82	0.82
VWC	0.85	0.82	1.00	0.83
SFC	0.80	0.82	0.83	1.00

Table 6: Domain Score Correlations: Papers Flagged

To examine the effect of the formatting issue on schools and districts, we plotted the mean total writing score by the percent of papers flagged. In figure 3, we see that there is a slight increase in writing total as the percent of the percent of papers flagged increases. We can conservatively say that there is no evidence that having a larger percentage of students experiencing the formatting issue results in lower writing scores.

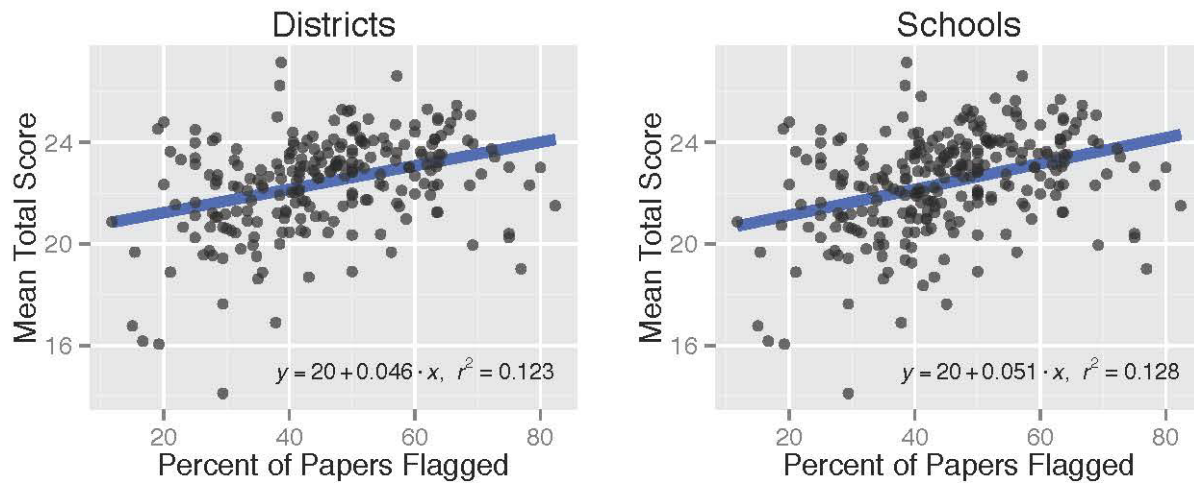


Figure 3: Grade 8 Writing Mean by Percent of Papers Flagged

Grade 11

The same analyses were repeated for grade 11, except that there are no grade 10 reading results available to do an ANCOVA. We see the same results at grade 11 with higher means for papers flagged for formatting problems. The Chi Squared test for the equality of correlation matrices (tables 9 and 10) is again significant 1872.341 (probability ≤ 0) with lower correlations for non-flagged papers and the same patterns of correlations. We also see in figure 4 that percent of papers flagged is not associated with lower writing scores at grade 11.

flag	Grade	IC	ORG	VWC	SFC	Total
FALSE	11	5.63	5.63	5.68	5.56	22.51
TRUE	11	6.07	6.01	6.06	5.90	24.08

Table 7: Writing Score Means by Format Flag

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
flag	1	12557.75	12557.75	449.88	0.0000
Residuals	20819	581130.40	27.91		

Table 8: ANOVA:Writing Total by Format Flag

	IC	ORG	VWC	SFC
IC	1.00	0.92	0.92	0.88
ORG	0.92	1.00	0.91	0.88
VWC	0.92	0.91	1.00	0.92
SFC	0.88	0.88	0.92	1.00

Table 9: Domain Score Correlations: Papers Not Flagged

	IC	ORG	VWC	SFC
IC	1.00	0.85	0.86	0.79
ORG	0.85	1.00	0.82	0.78
VWC	0.86	0.82	1.00	0.85
SFC	0.79	0.78	0.85	1.00

Table 10: Domain Score Correlations: Papers Flagged

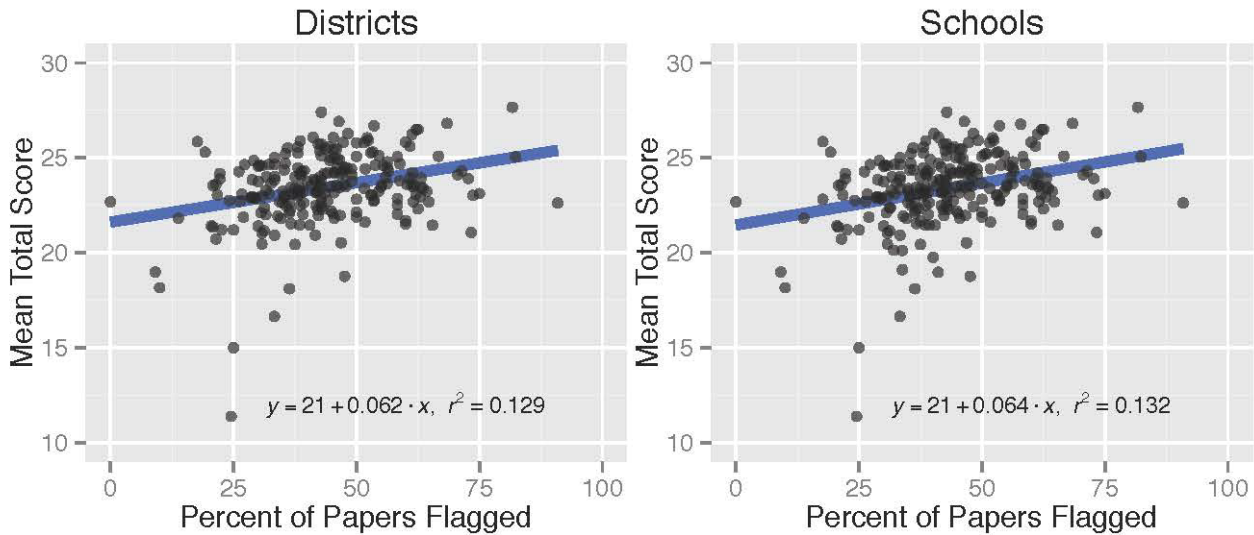


Figure 4: Grade 11 Writing Mean by Percent of Papers Flagged

Conclusion

While we can't conclude that the formatting issues had no negative effect on student scores, we don't have any evidence that student scores were systematically reduced. This does not reduce the seriousness of the problem or suggest that we can ignore the issue. We know from anecdotal reports by school staff that the formatting issues upset some students who experienced them. The statistical analyses reported here are limited by the data that are available. These results can only inform, not determine, interpretations of the scores or policy decisions to be made.

NeSA Writing 2013 - Supplemental Analysis

Nebraska NeSA Accountability TAC - April 24, 2013

The following analyses are supplemental information about the ODW formatting issue. In the original document, the graphs of Districts and Schools by Percent of Papers Flagged had kernel smoothed lines as a representation of the distributions. More traditional histograms with each bar representing a range of 5 percentage points are shown below.

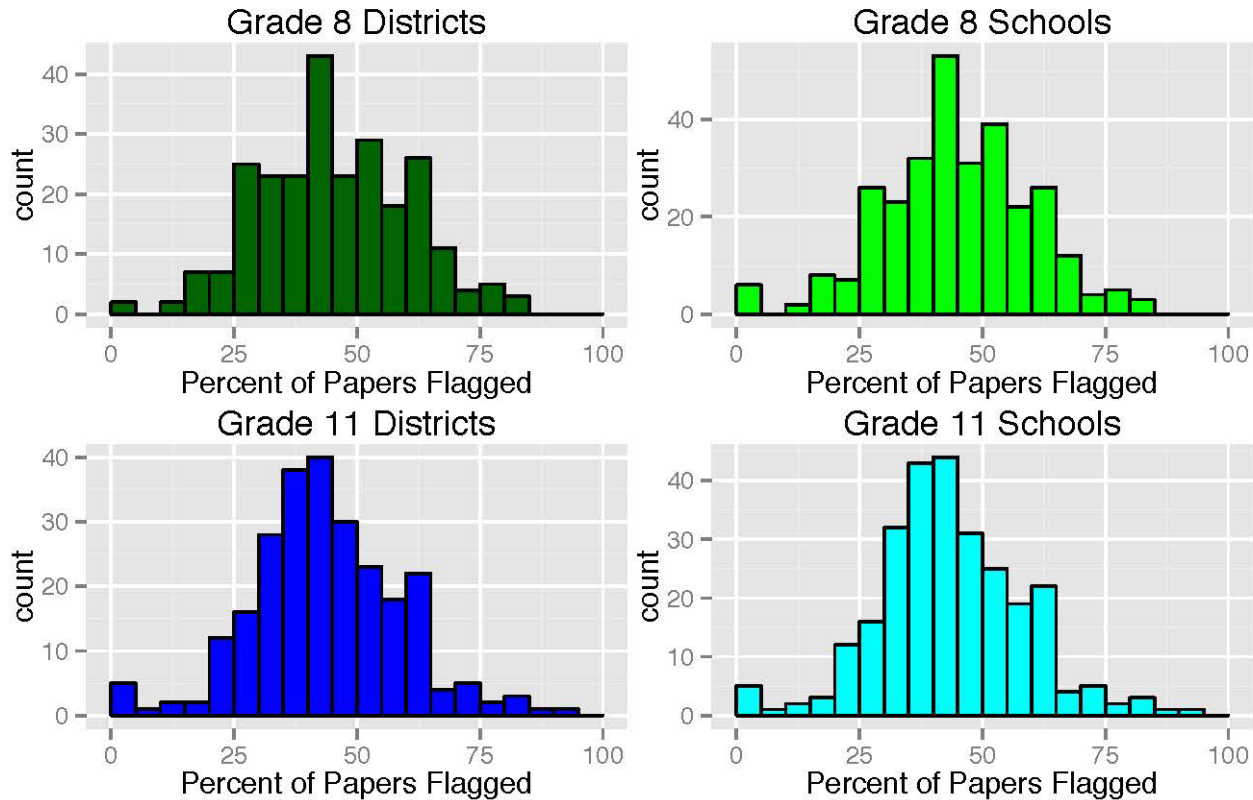


Figure 1: Histograms of Percent Flagged

We calculated the means, standard deviations (sd) and Cohen's d effect size (es) for the writing scores from flagged and non-flagged groups. The results are shown in Table 1.

grade	flag	func	IC	ORG	VWC	SFC	Total
08	FALSE	mean	5.58	5.50	5.64	5.52	22.24
		TRUE	5.85	5.71	5.88	5.68	23.14
	FALSE	sd	1.38	1.36	1.41	1.40	5.28
		TRUE	1.18	1.18	1.22	1.23	4.47
		es	0.21	0.16	0.18	0.12	0.18
11	FALSE	mean	5.63	5.63	5.68	5.56	22.51
		TRUE	6.07	6.01	6.06	5.90	24.08
	FALSE	sd	1.56	1.50	1.53	1.56	5.94
		TRUE	1.16	1.06	1.13	1.20	4.24
		es	0.31	0.28	0.28	0.24	0.30

Table 1: Writing Score Means, Standard Deviations and Effect Size by Format Flag

Limitations of the Research:

The most reliable method of identifying students who encountered formatting issues in the online engine would be tracking them through the engine software. However, that information is not available as the online engine did not track such details of test administration. Some students who encountered formatting issues easily ignored them or adjusted to them while other students experienced more frustration. This variation in students' experience cannot be accurately factored in when determining the effect of the formatting issues on student results. Because of these limitations, we cannot be certain about the extent of the impact on student results or school accountability.

Appendix L : Performance Assessment Services Formatting Response

Some students at grades 8 and 11 experienced formatting issues with the NeSA Writing online test administration. While research into score results does not indicate an effect on students' results, DRC undertook the following action prior to scoring in order to ensure accurate scoring and attempt to quantify the number of students impacted.

- The Scoring staff met with the Project Management team to review the list of issues reported by the districts.
- For any issues that would be visible to readers, the Scoring team selected examples from the pool of returned answer documents.
- Using these example papers, the Scoring team assembled a set for discussion. Prior to scoring, readers were trained to recognize the issues and instructed to ensure that these formatting issues would not factor into the scoring of the responses.
- Once acclimated to the 2013 formatting issues in the grade 8 and 11 online administrations, readers were instructed to indicate for each response scored, the possible presence or absence of formatting issues for information gathering purposes only.
- Throughout scoring, Team Leaders and Scoring Directors reviewed readers' work to ensure that scoring rubrics were being applied accurately for all responses regardless of formatting issues.