

Technical Report

Nebraska State Accountability (NeSA) Spring 2016 Writing Test

Grades 4, 8, and 11

October 2016



TABLE OF CONTENTS

1. GENERAL INFORMATION	
1.1 HISTORY	1
1.2 OVERVIEW	2
2. ADMINISTRATION OF THE WRITING ASSESSMENT	
2.1 WRITING TOPICS	3
2.2 TEST SESSIONS, TIMING, AND FORMAT	3
2.3 SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS	4
2.4 MATERIALS RETURN	5
2.5 TEST SECURITY MEASURES	5
2.6 SAMPLE MANUALS	5
3. PROCESSING AND SCORING THE NESA-WRITING	
3.1 RECEIPT OF MATERIALS	6
3.2 SCANNING OF MATERIALS	6
3.3 MATERIALS STORAGE	6
4. PERFORMANCE ASSESSMENT SERVICES (PAS)	
4.1 RANGEFINDING	8
4.2 TRAINING MATERIAL CREATION	9
4.3 READER RECRUITMENT/QUALIFICATIONS	9
4.4 TEAM LEADER AND READER TRAINING	9
4.5 HANDSCORING PROCESS	
4.6 QUALITY CONTROL	10
5. STUDENT DEMOGRAPHICS	13
6. REPORTING AND SCALING	15
7. RELIABILITY AND VALIDITY	
7.1 INTERNAL CONSISTENCY	
7.2 STANDARD ERROR OF MEASUREMENT	
7.3 INTER-RATER RELIABILITY	
7.4 DECISION CONSISTENCY AND ACCURACY	19
7.5 VALIDITY	
REFERENCES	21

APPENDIX A: NEBRASKA DEPARTMENT OF EDUCATION SCORING GUIDE FOR NARRATIVE WRITING – ANALYTIC – GRADE 4
APPENDIX B: NEBRASKA DEPARTMENT OF EDUCATION SCORING GUIDE FOR DESCRIPTIVE WRITING – ANALYTIC – GRADE 8
APPENDIX C: NEBRASKA DEPARTMENT OF EDUCATION SCORING GUIDE FOR PERSUASIVE WRITING – ANALYTIC – GRADE 11
APPENDIX D: PERFORMANCE LEVEL DESCRIPTORS GRADE 4 25
APPENDIX E: PERFORMANCE LEVEL DESCRIPTORS GRADE 8
APPENDIX F: PERFORMANCE LEVEL DESCRIPTORS GRADE 11 27
APPENDIX G: COMPOSITE TO SCALE SCORE TABLES GRADE 4
APPENDIX H: COMPOSITE TO SCALE SCORE TABLES GRADE 8
APPENDIX I: COMPOSITE TO SCALE SCORE TABLES GRADE 11

APPENDIX J: WRITING TECHN	OLOGY EVENTS SUMMARY.	34
---------------------------	-----------------------	----

1. GENERAL INFORMATION

1.1 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/FloorDocs/100/PDF/Final/LB1157.pdf) 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 1.1.1. 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.

Subject	Adminis	tration Year	Crades			
Subject	Field Test	Operational	Grades			
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			

Table 1.1.1 NeSA Administration Schedule

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 (NeSA-W) assessment under the direction of the Department of Education.

NeSA-W has been phased in as described in Table 1.1.2.

Table 112	NeSA-W	Administration	Schedule
1 aute 1.1.2	INCOA-W	Aummisuation	JUIEUUIE

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.

1.2 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 was the provider of the printed and online versions of the 2016 NeSA-W Tests.

Paper/Pencil and Online Testing Window: January 18 – February 5, 2016 Number of Potential Testing Sites 250 districts 938 schools

2016 Testing Outages

Due to technology difficulties experienced by some students at grades 8 and 11 in Nebraska, the NDE Assessment and Accountability team requested that the Nebraska Technical Advisory Committee review the description of the technical issues and the effect of the technology issues on the NeSA-Writing score results. The TAC determined that the NeSA-Writing scores for grades 8 and 11 on NeSA-Writing are valid and reliable and could be released for reporting and used in accountability. The Writing Technology Events Summary document can be found in Appendix J.

2. Administration of the Writing Assessment

2.1 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

2.2 TEST SESSIONS, TIMING, AND FORMAT

The test window for the grade 4 paper/pencil tests, including make-up tests, was January 18 – February 5, 2016. 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 18 – February 5, 2016. 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-W paper/pencil test as well as the grades 8 and 11 NeSA-W online tests were available to all schools. Spanish versions of the tests were made available to all districts. Table 2.2.1 shows the number of student who took each exam by mode of administration.

Grade	Number of Students Tested Paper/Pencil	Number of Students Tested Online
4	23,075	N/A
8	514	21,732
11	439	21,031

Table 2.2.1 2016 NeSA-W Test Participation

Tables 2.2.2 and 2.2.3 contain 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 unlocked 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.

Time Span in	Student Count	% in Each Time
Minutes		Span
0-10	91	.42
10-20	130	.60
20-30	267	1.23
30-40	655	3.01
40-50	1,156	5.32
50-60	1,750	8.05
60-70	2,418	11.13
70-80	2,750	12.65
80-90	2,209	10.16
90+	10,306	47.42
Total	21,732	100.00

Table 2.2.2 2016 NeSA-W Grade 8 Online Test Times

Table 2.2.3 2016 NeSA-W Grade 11 Online Test Times

Time Span in	Student Count	% in Each Time
Minutes		Span
0-10	62	.29
10-20	228	1.08
20-30	710	3.38
30-40	1,675	7.96
40-50	2,732	12.99
50-60	3,149	14.97
60-70	3,140	14.93
70-80	2,632	12.51
80-90	1,806	8.59
90+	4,897	23.28
Total	21,031	100.00

2.3 SHIPPING, PACKAGING, AND DELIVERY OF MATERIALS

A single shipment was sent out by DRC to each district. The shipment was delivered by January 4, 2016. The shipment contained all necessary materials to complete the NeSA-W 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 32,197 standard writing booklets, 504 Spanish translation booklets, 3,072 manuals, and approximately 7,500 non-secure materials for testing sites. DRC used UPS to deliver materials to the testing sites. DRC used UPS to deliver materials to the testing sites.

2.4 MATERIALS RETURN

The materials return window was February 10-12, 2016. DRC used UPS for all return shipments.

2.5 TEST SECURITY MEASURES

Test security is essential to obtaining reliable and valid scores for accountability purposes. The 2016 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 October 30, 2015. 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.

2.6 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 <u>www.education.ne.gov/assessment</u>.

3. PROCESSING AND SCORING THE NeSA-WRITING

3.1 RECEIPT OF MATERIALS

Receipt of NeSA-Writing materials began on February 10, 2016, and concluded on February 22, 2016. Any materials received after February 22, 2016, were considered late and were checked-in, scanned, and processed during the late window of February 24, 2016 through March 25, 2016. 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.

3.2 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.

3.3 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.

4. Performance Assessment Services

In 2016, the Nebraska Department of Education (NDE) continued the use of analytic scoring rubrics for grades 4, 8 and 11. 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.

4.1 RANGEFINDING

After receiving student responses from the 2015 NeSA-W Field Test, DRC's Performance Assessment Services (PAS) staff reviewed responses for the chosen 2016 operational prompts for each grade (4, 8, and 11) and assembled them into item-specific sets, each with 120 responses that exemplified the range of possible score points.

When selecting responses for the rangefinding sets, care was taken to include a number of responses that potentially fall on the edges of a score range. These responses are specifically selected by PAS staff in order to identify where the different score points begin and end. Additionally, responses that potentially helped exemplify the differences between domains were selected (i.e., responses that potentially received different scores in different domains). Copies of these sets were made for each member of the rangefinding committee. DRC's PAS staff then travelled to Lincoln, Nebraska, (June 17-18, 2015) and facilitated the rangefinding sessions.

The rangefinding committees consisted of Nebraska educators and NDE staff members. 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 gradespecific committees consisting of ten or twelve Nebraska educators and an NDE representative. Each committee reviewed the current prompt, scoring rubric, and the Scoring Guide anchor papers from the spring 2015 NeSA-W Operational Test. Following this review and discussion, each committee then began to consensus score one of the field test items.

Initially, each student response was read aloud and then discussed by all members of the group to ensure that everyone was interpreting the analytic rubric consistently and uniformly. Each of the four domains was addressed independently for each response, 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 for each field test item in their grade.

Discussions of student responses relied on the 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 as the committee members described and justified scoring decisions. This information was used by the Scoring Directors and Team Leaders during reader training.

4.2 TRAINING MATERIAL CREATION

As part of preparation for the 2016 NeSA-Writing assessment scoring, DRC's PAS staff assembled the committee scored rangefinding responses into sets used for training readers. Responses that the rangefinding committee selected as relevant in terms of the scoring concepts they illustrated were annotated and included as anchor papers in a Scoring Guide. The 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 remaining student responses that were reviewed and scored by rangefinding committee members. Responses were selected for training to show readers the spectrum for each score point in each domain and to highlight some of the writing characteristics within each domain. Additionally, an effort was made to include responses that received different scores across the domains in order to ensure that scorers clearly understood the differences between each of the domains.

4.3 READER RECRUITMENT/QUALIFICATIONS

The Scoring Director and Team Leaders were chosen by the Project Manager from a pool, consisting of experienced individuals who are proven successful readers and supervisors with strong backgrounds in Nebraska writing. The selected staff demonstrated organization, leadership, and management skills. DRC retains a pool of experienced readers from year to year and all of the 2016 NeSA-Writing readers came from this population. All scoring personnel were required to sign confidentiality agreements before any training or handling of secure materials began. It should be noted that the Scoring Directors for this project have remained constant since 2014 for grade 4 and 2011 for grades 8 and 11.

4.4 TEAM LEADER AND READER TRAINING

Representatives from NDE were on site at the Plymouth, Minnesota Scoring Center (February 4-12, 2016) to collaborate with DRC Scoring Directors and Team Leaders during three-day training sessions. The scoring Project Manager, Scoring Director, thirteen Team Leaders, and a representative from NDE worked cooperatively to review and discuss all of the training materials, and to consensus score a number of additional validity papers for each grade.

Two days of reader training took place on February 9-10, 2016, for grades 8 and 11, and February 16-17, 2016, for grade 4. 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 a small-group setting.

Once the scoring rubric, anchor sets, and training sets were thoroughly discussed, each reader 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 before proceeding to subsequent qualifying sets. Readers were required to achieve 70% exact agreement in each domain on at least of the qualifying sets in order to stay on the project and score any live student responses. In 2016, 56 readers were qualified to score Nebraska grade 4 student writing responses, 44 readers were qualified to score Nebraska grade 8 student writing responses, and 45 readers were qualified to score Nebraska grade 11 student writing responses.

Following training and qualifying, a period of paired scoring took place, during which pairs of readers were required to work cooperatively to score live responses by discussing and agreeing on the appropriate score. Once Team Leaders were satisfied with their performance, the readers were permitted to score independently while being monitored closely.

4.5 HANDSCORING PROCESS

DRC handscoring system, ScoreBoard, automatically routes student responses to qualified readers until all required first, second, and adjudication reads have been completed. Readers cannot tell if they are the first or second reader. Nebraska 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 response was scored twice, and non-adjacent scores were adjudicated. Data collected from the multiple reads was used to calculate the inter-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 scoring Project Manager 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.

4.6 QUALITY CONTROL

Validity sets

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.

Feedback from the scoring of validity responses provides a great deal of information as to how accurately and consistently readers are scoring over the course of the 6 to 8 days of

scoring. Cumulative summary information from the Grade 4, 8, and 11 Validity Item Detail Report is shown in Table 4.6.1 below as an example of this data.

GRADE	IDEAS/CONTENT			ORO	ORGANIZATION			VOICE/WORD CHOICE			SENTENCE FLUENCY/CONVENTIONS			
Match	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ		
4	76%	24%	100%	78%	22%	100%	78%	22%	100%	80%	20%	100%		
8	85%	15%	100%	80%	20%	100%	82%	17%	99%	71%	27%	98%		
11	86%	14%	100%	87%	13%	100%	83%	17%	100%	82%	18%	100%		

Table 4.6.1 Validity Set Reader Agreement for NeSA-W 2016

Recalibration Sets

During the course of scoring, two recalibration sets were produced using pre-determined scored student responses. These sets were 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 Inter-Rater Reliability Report (Table 4.6.2) and Score Point Distribution Report (Table 4.6.3) were especially helpful in analyzing scoring data and maintaining high standards of scoring quality.

GRADE	IDEAS/CONTENT			ORG	ORGANIZATION			VOICE/WORD CHOICE			SENTENCE FLUENCY/CONVENTIONS			
Match	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ	EXACT	ADJ	EX +ADJ		
4	70%	29%	99%	69%	30%	99%	68%	31%	99%	69%	30%	99%		
8	74%	26%	100%	73%	26%	99%	73%	27%	100%	72%	28%	100%		
11	73%	27%	100%	74%	26%	100%	72%	28%	100%	70%	29%	99%		

Table 4.6.2 Inter-rater Reliability Results for NeSA-W 2016

	Table 4.6.3 Score Point Distributions for NeSA-W 2016															
GRADE	E IDEAS/CONTENT					RADE IDEAS/CONTENT ORGANIZATION VOICE				CE/WO	ORD CH	OICE	FLUE	SENT NCY/C	TENCE ONVEN	TIONS
Score Point	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
4	2%	33%	52%	9%	3%	34%	51%	9%	3%	34%	50%	10%	4%	33%	49%	10%
8	1%	22%	58%	17%	2%	23%	58%	16%	2%	22%	57%	18%	2%	25%	56%	17%
11	2%	17%	61%	19%	3%	13%	65%	18%	2%	16%	62%	19%	3%	18%	60%	18%

Table 4.6.3 Score Point Distributions for NeSA-W 2016

5. STUDENT DEMOGRAPHICS

Gender, ethnicity, free or reduced lunch status (FRL), Limited English Proficiency/English Language Learners (LEP/ELL) status, Special Education status (SPED), and accommodation status data were collected for all students who participated and attempted the 2016 NeSA-Writing assessments. This summary of student demographics by grade is provided in Table 5.1. The table shows that for each grade, over 21,000 students took the assessment. Of those students across grades, half are males, half are females, over half are white, and less than one fifth are Hispanic. Among the students across grades, about 37% to 46% are eligible for FRL, 2% to 7% are LEP/ELL, and 10% to 15% belong to at least one SPED category. For all three of these programs/categories, the participation rate is lower for upper grade students. In terms of the test accommodations, there are about 6% to 16% of the students across grade and content area that report at least one type of accommodation (see row 'Total' for 'Accommodation' in the table). Similar to the rate for FRL, LEP/ELL, and SPED across grades, the rate for accommodation is lower for high school students (Grade 11). Across all grades, the 'Timing/Schedule/Setting' is the most utilized accommodation (about 6-9% for Grade 4 and 8, and 4% for Grade 11), followed by the 'Indirect Linguistic Support in grade 4 (6%) and for grades 8 and 11, 'Response' (about 3-5%).

		Grade 4		Grad	le 8	Grade 11	
		Count	%	Count	%	Count	%
All Students		23075	100.0	22246	100.0	21470	100.0
Gender	Female	11300	49.0	10926	49.1	10420	48.5
	Male	11775	51.0	11320	50.9	11049	51.5
	American Indian/Alaska Native	327	1.4	280	1.3	280	1.3
	Asian	583	2.5	512	2.3	536	2.5
	Black	1600	6.9	1387	6.2	1268	5.9
Race/Ethnicity	Hispanic	4321	18.7	4002	18.0	3562	16.6
	Native Hawaiian or other Pacific Islander	31	0.1	27	0.1	33	0.2
	White	15365	66.6	15311	68.8	15117	70.4
	Two or More Races	848	3.7	727	3.3	673	3.1
FRL	Yes	10661	46.2	9413	42.3	7863	36.6

Table 5.1 NeSA-W Summary data: Demographics and Accommodations

		Grad	le 4	Grade 8		Grade 11	
		Count	%	Count	%	Count	%
	No	12101	52.4	12541	56.4	13240	61.7
LEP/ELL	Yes	1630	7.1	570	2.6	499	2.3
	No	21445	92.9	21676	Grade 8Grade 11Count%Count%1254156.41324065702.649992167697.4209719297613.4223311927086.61923788043.6333 $-$ 11475.2544 $-$ 14116.3836 $-$ 2901.3162 $-$ 2671.2158 $-$ 1000.478 $-$ 10.00 $-$ 20249.11233 $-$	97.7	
Special	Yes	3505	15.2	2976	13.4	2233	10.4
Education	No	19570	84.8	19270	86.6	19237	89.6
	Content Presentation	712	3.1	804	3.6	333	1.6
	Response	918	4.0	1147	5.2	544	2.5
	Timing/Schedule/Setting	2169	9.4	1411	6.3	836	3.9
	Direct Linguistic Support with Test Directions	805	3.5	290	1.3	162	0.8
Accommo- dations	Direct Linguistic Support with Content and Test items	649	2.8	267	1.2	158	0.7
	Indirect Linguistic Support	1429	6.2	166	0.7	93	0.4
	Spanish	82	0.4	100	0.4	78	0.4
	Braille*	3	0.0	1	0.0	0	0.0
	Large Print*	1	0.0	2	0.0	4	0.0
	Total	3682	16.0	2024	9.1	1233	5.7

*Count represents the number of booklets ordered. This is not tracked.

6. REPORTING AND SCALING

As mentioned in Chapter 4, the Nebraska writing scoring rubric uses one prompt, four domains, and two readers with scores of 1 to 4. If all scores were simply added up the result would be 25 discrete score points ranging from 8 to 32.

To create an equal-interval scale for the NeSA-W, scale scores were assigned to each raw score point through a linear transformation of the logit scores. Scale scores do not alter the relationships or the displays. Scale scores are the numbers that will be reported to describe the performance of the students, schools, and systems. They will define the ranges of the performance levels, appear on individual student reports and school accountability analyses, and be dissected in newspaper accounts.

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 Rasch logit metric, when it can be used, has better measurement properties than any linear transformation of raw scores.

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

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 2016 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	Composite score
Reader 1	3	3	2	3	11.2
	(4.2)	(3)	(1.6)	(2.4)	11.2
Reader 2	3	2	3	3	11.0
	(4.2)	(2)	(2.4)	(2.4)	11.0

*Note: Weighted calculations are in parentheses.

The 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*. A summary of the frequency distributions of the state scale scores for the NeSA-W is provided in Table 6.1.1.

		Scale Score		Perce	ntile Scale Score	
Grade	Count	Mean	S.D.	25%	50%	75%
4	23075	43.4	12.6	36	43	50
8	22246	46.1	13.4	37	48	55
11	21470	45.9	13.6	38	46	56

Table 6.1.1 2016 NeSA-W State Scale Score Summary, All Students

As part of its deliberations concerning defining the performance levels, the State Board of Education specified that the *Meets the Standards* performance level have a scale score of 40 for all grades and that the *Exceeds the Standards* level have a scale score of 57 for Grade 4, 55 for Grade 8, and 53 for Grade 11. The standards defining the performance levels were adopted by the SBE per the standard setting and standard validation completed in 2012 for Grade 8 and 11, and in 2013 for Grade 4. Complete documentation of all standard setting events are presented in separate documents labeled *NeSA Spring 2012 Writing Test Technical Report*, and *NeSA Spring 2013 Writing Test Technical Report*, which may be found on the Nebraska State Department of Education website. Note that the scale score values that define the performance levels are fixed and will not change from year to year. The percentage of Spring 2016 students in each performance level are shown below in Table 6.1.2.

le 0.1.2 2010 NeSA-W State I erformance Level Summary, An Stud									
	Below		Meet		Exceed				
Grade	Count	Percent	Count	Percent	Count	Percent			
4	6936	30.1	12797	55.5	3342	14.5			
8	6562	29.5	9790	44.0	5894	26.5			
11	5989	27.9	9307	43.3	6174	28.8			

Table 6.1.2 2016 NeSA-W State Performance Level Summary, All Students

DRC reported student results on the NeSA-W 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.

7. RELIABILITY AND VALIDITY

This chapter addresses the reliability and validity of the NeSA-W test scores. Reliability refers to the degree to which test scores are consistent over repeated measurements and validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests, according to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014).

7.1 INTERNAL CONSISTENCY

The ability to measure consistently is a necessary prerequisite for making appropriate interpretations (i.e., showing evidence of valid use of results). Conceptually, reliability can be referred to as the consistency of the results between two measures of the same thing. This consistency can be seen in the degree of agreement between two measures on two occasions. Operationally, such comparisons are the essence of the mathematically defined reliability indices.

One important reliability index is when we use a single measurement instrument, administered to a group of people on one occasion, to estimate reliability. In effect we judge the reliability of the instrument by estimating how well the items that reflect the same construct yield similar results. Thus, we investigate how consistent the results are for different items for the same construct within the measure.

Given the one-prompt, four-domain structure of the NeSA-W test, it is interesting to see how performance on one domain correlates to that on the other domains. One index that directly assesses the extent to which answers to one domain correlate with answers to other domains is the average inter-domain correlation. For a shorter test, the reliability index of the average inter-item (i.e., inter-domain) correlations is particularly important. Tables 7.1.1 to 7.1.4 report the inter-domain correlations for Grade 4, 8, and 11. The reliability of the average inter-domain correlations is presented in Table 7.1.5 for each grade.

Code Domain							
D.1	Ideas/Content						
D.2	Organization						
D.3	Voice/Word Choice						
D.4	Sentence Fluency/Conventions						

Table 7.1.1 NeSA-W Domains

Grade 4	D.1	D.2	D.3	D.4
D.1				
D.2	0.95			
D.3	0.92	0.92		
D.4	0.89	0.89	0.92	

Table 7.1.2 Correlations between Domain Scores: Grade 4

Table 7.1.3 Correlations between Domain Scores: Grade 8

Grade 4	D.1	D.2	D.3	D.4
D.1				
D.2	0.92			
D.3	0.92	0.91		
D.4	0.86	0.87	0.88	

Table 7.1.4 Correlations between Domain Scores: Grade 11

Grade 4	D.1	D.2	D.3	D.4
D.1				
D.2	0.92			
D.3	0.93	0.90		
D.4	0.86	0.85	0.90	

Table 7.1.5 Form Reliability

Grade	Reliability
4	0.92
8	0.89
11	0.89

7.2 STANDARD ERROR OF MEASUREMENT

The Rasch model, which is used to analyze the writing assessment, provides asymptotic standard errors for each raw score. These standard errors are often referred to as conditional standard errors (CSEM) (Wright & Masters, 1982) to differentiate them from the standard error that is often used in the true-score model. These asymptotic standard errors for each raw score can be found in Appendix G, H and I. The CSEMs are presented in the scale score metric.

7.3 INTER-RATER RELIABILITY

Because the scoring of the writing tasks involves at least two independent readers, another source of random error is related to the variation across readers in the measurement procedures and interpretation of measurement results. To address these sources of error variance in rating measurements for reliability, DRC's Performance Assessment Services (PAS) follows a series of strict procedures in reader recruitment, reader training, and validity control, as is detailed in Chapter 4. As a result, the degree of agreement among raters, as provided in Table 4.6.1, is acceptable at about 75% exact agreement rate.

Further inter-rater reliability information is provided by the implementation of validity set as one of the PAS quality control procedures in scoring. As discussed in Chapter 4, the validity set is pre-scored responses helped to track consistency over time and how well individual reader were performing. As reported in Table 4.6.1 the exact agreement rate between readers on the validity set is approximately 80% on average.

7.4 DECISION CONSISTENCY AND ACCURACY

When criterion-referenced tests are used to place the examinees into two or more performance classifications, it is useful to have some indication of how accurate or consistent such classifications are. Decision consistency refers to the degree to which the achievement level for each student can be replicated upon retesting using an equivalent form (Huynh, 1976). Decision accuracy describes the extent to which achievement-level classification decisions based on the administered test form would agree with the decisions that would be made on the basis of a perfectly reliable test. In a standards-based testing program, there should be great interest in knowing how consistently and accurately students are classified into performance categories. Since it is not feasible to repeat NeSA testing in order to estimate the proportion of students who would be reclassified in the same achievement levels, a statistical model needs to be imposed on the data to project the consistency or accuracy 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. These approaches are fairly complex, and the cited sources contain details regarding the statistical models used to calculate decision consistency from the single NeSA administration.

Several factors might affect decision consistency. One important factor is the reliability of the scores. All other things being equal, more reliable test scores tend to result in more similar reclassifications. Another factor is the location of the cutscore in the score distribution. More consistent classifications are observed when the cutscores are located away from the mass of the score distribution. The number of performance levels is also a consideration. Consistency indices for four performance levels should be lower than those based on three categories because classification using four levels would allow more opportunity to change achievement levels. Finally, some research has found that results from the Hanson and Brennan (1990) method on a dichotomized version of a complex assessment yield similar results to the Livingston and Lewis method (1995) and the method by Stearns and Smith (2007).

The results for the overall consistency across all three achievement levels are presented in Table 7.4.1. The tabled values, derived using the program *BB-Class* (Brennan, 2004), show that consistency values across the two methods are generally very similar. Across all grades, the overall decision consistency ranged from the mid 0.80s to the low 0.90s while the decision accuracy ranged in the low to mid 0.90s. If a parallel test were administered, at least 85% or more of students would be classified in the same way. Dichotomous decisions using the Meets

cuts generally have the slightly higher consistency values and exceeded 0.90 in most cases. The pattern of decision accuracy across different cuts is similar to that of decision consistency.

Contont	Livingston & Lewis			Hanson & Brennan					
Area	Grade	Decision Accuracy		Decision Consistency		Decision Accuracy		Decision Consistency	
		Meets	Exceeds	Meets	Exceeds	Meets	Exceeds	Meets	Exceeds
Writing	4	0.92	0.94	0.88	0.92	0.92	0.94	0.89	0.92
	8	0.91	0.92	0.88	0.89	0.93	0.94	0.90	0.91
	11	0.91	0.91	0.88	0.88	0.93	0.93	0.90	0.90

Table 7.4.1 NeSA-W	Decision C	Consistency	Results

7.5 VALIDITY

Content validity addresses whether the test adequately samples the relevant material it purports to cover. The NeSA-W for grades 4, 8, and 11 is a criterion-referenced assessment. The criteria referenced are the Nebraska writing content standards. The assessment was based on, and was directly aligned to, the Nebraska statewide content standards to ensure good content validity.

For criterion-referenced, standards-based assessment, strong content validity evidence is derived directly from the test construction process and the item scaling. The item development and test construction process, described above, ensures that every item aligns directly to one of the content standards. This alignment is foremost in the minds of the item writers and editors. As a routine part of item selection and prior to an item appearing on a test form, the review committees check the alignment of the items with the standards and make any adjustments deemed necessary. The result is a mutual agreement among the content specialists and teachers that the assessment does in fact assess what was intended.

Evidence of this agreement is reflected in the success of the Body of Work standard setting processes (in the separate *NeSA Spring 2012 Writing Test Technical Report*, and *NeSA Spring 2013 Writing Test Technical Report*). Panelists participating in the Body of Work process read a sample of essays in a wide range from very low to very high levels. Discussions about placement of each individual essay almost invariably focus on the knowledge, skills, and behaviors required of a typical student in each grade, and, overall, panelists were comfortable with the content coverage of each writing task.

As described in the *Standards* (2014), internal-structure evidence refers to the degree to which the relationships between test items and test components conform to the construct on which the proposed test interpretations are based. As discussed in Section 7.1, the inter-domain correlations are all positive and of acceptable magnitude. This also provided evidence that the four domains were essentially unidimensional, and supported the interpretations based on the total composite scores for the NeSA-W test.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Brennan, R. L., & Hanson, B. A. (2004). BB-Class (Version 1.0). [Computer software] Iowa City, IA: University of Iowa, Center for Advanced Studies in Measurement & Assessment. Available at CASMA: education.uiowa.edu/casma.
- Hambleton, R., & Novick, M. (1973). Toward an integration of theory and method for criterionreferenced tests. *Journal of Educational Measurement*, *10*, 159–170.
- Hanson, B. A., & Brennan, R. L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score theory models. *Journal of Educational Measurement*, 27, 345–359.
- Huynh, H. (1976). On the reliability of decisions in domain-referenced testing. *Journal of Educational Measurement*, 13, 253–264.
- Livingston, S., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement 32*, 179–197.
- Stearns, M., & Smith R. M. (2008). Estimation of classification consistency indices for complex assessments: Model based approaches. *Journal of Applied Measurement*, *9*, 305-315.
- Wright, B. D., & Masters, G. N. (1982) Rating scale analysis. Chicago, IL: MESA Press.

	Nebraska Department o	of Education Scoring Guide	for Narrative Writing – Ar	nalytic – Grade 4
	1	2	3	4
IDEAS / CONTENT 35%	 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. 	 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. 	 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. 	 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%	 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. 	 Structural development of a beginning, middle, or end is limited. Pacing is somewhat inconsistent. Transitions are repetitious or weak. Paragraphing is irregular. 	 Structural development of a beginning, middle, and end is functional. Pacing is generally controlled. Transitions are functional. Paragraphing is generally successful. 	 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%	 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. 	 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. 	 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. 	 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%	 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. 	 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. 	 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. 	 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 A: Nebraska Department of Education Scoring Guide for Narrative Writing – Analytic – GRADE 4

Nebr	aska Department of Edu	cation Scoring Guide for	Descriptive Writing - A	nalytic - GRADE 8
	1	2	3	4
IDEAS / CONTENT 35%	 The picture of what is being described is unclear. Content has many digressions from the topic. Sensory details are lacking. 	 The picture of what is being described is limited. Content has some digressions from the topic. Sensory details are limited or unrelated. 	 The picture of what is being described is clear. Content is generally focused on the topic. Sensory details are adequate and related. 	 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%	 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. 	 Structural development of an introduction, body, and conclusion is limited. Pacing is somewhat inconsistent. Transitions are repetitious or weak. Paragraphing is irregular. 	 Structural development of an introduction, body, and conclusion is functional. Pacing is generally controlled. Transitions are functional. Paragraphing is generally successful. 	 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%	 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. 	 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. 	 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. 	 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%	 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. 	 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. 	 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. 	 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

Appendix C: Nebra	ska Department of E	ducation Scoring	Guide for Persuasive	Writing – Analytic – GRADE 1
-------------------	---------------------	------------------	----------------------	------------------------------

Nebr	raska Department of Edu	cation Scoring Guide for	Persuasive Writing – An	alytic - GRADE 11
	1	2	3	4
IDEAS / CONTENT 35%	 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. 	 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. 	 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. 	 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%	 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. 	 Structural development of an introduction, body, and conclusion is limited. Pacing is somewhat inconsistent. Transitions are repetitious or weak. Paragraphing is irregular. 	 Structural development of an introduction, body, and conclusion is functional. Pacing is generally controlled. Transitions are functional. Paragraphing is generally successful. 	 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%	 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. 	 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. 	 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. 	 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%	 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. 	 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. 	 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. 	 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										
Below the Standards	Meets the Standards	Exceeds 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.	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.	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 is below the standards if the	The student's writing meets the standards if the	The student's writing exceeds the standards if the								
 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. 	 Writer creates a general understanding of events in the story. Content is generally focused on the topic. Details are adequate and related. Storyline is generally logical and easy to follow. Structural development of a beginning, middle, and end 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. Sentences 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. 	 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 maninulated for 								

Below the Standards Meets the Standards Exceeds 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 reflects a satisfactory performance of the standards and a sufficient understanding of the traits of writing. The student's writing is necessary. Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the traits of writing. The student's writing demonstrates more strengths than weaknesses. Some revision and/or editing is necessary. Overall the student's writing demonstrates more of the standards if the Overall the student's writing demonstrates more strengths than weaknesses. Some revision and/or editing is necessary. Overall the student's writing exceeds the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the traits of writing. The student's writing meets the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understates numerous necessary. The student's writing is below the standards if the The student's writing meets the standards if the Or that is being described is clear an obdy, and conclusion is limited or indexing. The student's writing neerelity standards if the Or anarstons are repetitions, week, unclear, or missing. Overall the student's writing neerelity standards if the Or Paragraphing is generally sports and functional. Or anarstons are repetitions, week, unclear, or missing. Overall the student's writing neerelithopublic. Or anarstons are repetitons, week, unclear,	Nebraska State Accou	Intability-Writing (NeSA-W) Performa	nce Level Descriptors
Below the Standards Meets the Standards Overall the student's writing reflects an unsatisfactory performance of the standards and a nusficient understanding of the traits of writing. The student's writing reflects as atisfactory performance of the standards and a sufficient understanding of the traits of writing. The student's writing reflects as atisfactory performance of the standards and a thorough understanding of the straits of writing. The student's writing reflects as atisfactory performance of the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the writing. The student's writing reflects as atisfactory performance of the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the writing. The student's writing reflects as atisfactory performance of the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the writing. The student's writing reflects as atisfactory performance of the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the writing. The student's writing reflects as atisfactory performance of the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the traits of writing. The student's writing reflects as atisfactory performance of the standards if the Overall the student's writing reflects an advanced per of the standards and a thorough understanding of the traits of writing. The student's writing reflects as real advanced per of the standards and a thorough understanding of the traits of writing. Exceeds the standards		Grade 8	
 Grammar, usage, punctuation, and spelling stylistic effect. 	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 • Picture of what is being described is limited or unclear. • Content has some digressions from the topic. • Sensory details are limited, unrelated, or lacking. • Pacing is inconsistent or awkward. • Transitions are repetitious, weak, unclear, or missing. • Paragraphing is irregular, ineffective, or missing. • 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.	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 • Picture of what is being described is clear. • Ontent 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 appropriate for the purpose and audience. • Language is generally specific, precise, and varied. • Writing has adequate vivid words and phrases. • Sentences 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.	 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 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. 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
errors distract the reader.	 the reader. Grammar, usage, punctuation, and spelling errors distract the reader. 	usually correct and rarely distract the reader.	consistently correct and may be manipulated for stylistic effect.

Appendix F: Performance Level Descriptors Grade 11

Nebraska State Accou	intability-Writing (NeSA-W) Performa Grade 11	nce Level Descriptors
 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. Paring is generally controlled. Transitions are functional. Paragraphing is generally successful. Writer demonstrates a general commitment to the topic. Language is generally specific, precise, varied, and engaging. Writer generally anticipates the reader's questions. Sentences 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.

	Composite	Scale			Composite	Scale			Composite	Scale	
Grade	Score	Score	CSEM	Grade	Score	Score	CSEM	Grade	Score	Score	CSEM
4	8.0	1	8	4	12.0	19	2	4	16.0	29	5
4	8.1	7	8	4	12.1	19	2	4	16.1	31	5
4	8.2	10	5	4	12.2	19	2	4	16.2	32	4
4	8.3	11	4	4	12.3	19	2	4	16.3	32	4
4	8.4	12	4	4	12.4	20	2	4	16.4	33	4
4	8.5	13	3	4	12.5	20	2	4	16.5	33	3
4	8.6	13	3	4	12.6	20	2	4	16.6	34	3
4	8.7	13	3	4	12.7	20	2	4	16.7	34	3
4	8.8	14	3	4	12.8	20	2	4	16.8	35	3
4	8.9	14	3	4	12.9	20	2	4	16.9	35	3
4	9.0	14	3	4	13.0	20	2	4	17.0	35	3
4	9.1	15	2	4	13.1	20	2	4	17.1	35	2
4	9.2	15	2	4	13.2	21	2	4	17.2	36	2
4	9.3	15	2	4	13.3	21	2	4	17.3	36	2
4	9.4	15	2	4	13.4	21	2	4	17.4	36	2
4	9.5	16	2	4	13.5	21	2	4	17.5	36	2
4	9.6	16	2	4	13.6	21	2	4	17.6	36	2
4	9.7	16	2	4	13.7	21	2	4	17.7	37	2
4	9.8	16	2	4	13.8	21	2	4	17.8	37	2
4	9.9	16	2	4	13.9	22	2	4	17.9	37	2
4	10.0	16	2	4	14.0	22	2	4	18.0	37	2
4	10.1	17	2	4	14.1	22	2	4	18.1	37	2
4	10.2	17	2	4	14.2	22	2	4	18.2	37	2
4	10.3	17	2	4	14.3	22	2	4	18.3	38	2
4	10.4	17	2	4	14.4	22	2	4	18.4	38	2
4	10.5	17	2	4	14.5	23	2	4	18.5	38	2
4	10.6	17	2	4	14.6	23	2	4	18.6	38	2
4	10.7	17	2	4	14.7	23	2	4	18.7	38	2
4	10.8	18	2	4	14.8	23	2	4	18.8	38	2
4	10.9	18	2	4	14.9	23	2	4	18.9	38	2
4	11.0	18	2	4	15.0	24	3	4	19.0	39	2
4	11.1	18	2	4	15.1	24	3	4	19.1	39	2
4	11.2	18	2	4	15.2	24	3	4	19.2	39	2
4	11.3	18	2	4	15.3	25	3	4	19.3	39	2
4	11.4	18	2	4	15.4	25	3	4	19.4	39	2
4	11.5	18	2	4	15.5	25	3	4	19.5	39	2
4	11.6	19	2	4	15.6	26	4	4	19.6	39	2
4	11.7	19	2	4	15.7	27	4	4	19.7	39	2
4	11.8	19	2	4	15.8	27	4	4	19.8	40	2
4	11.9	19	2	4	15.9	28	5	4	19.9	40	2

	Composite	Scale			Composite	Scale			Composite	Scale	
Grade	Score	Score	CSEM	Grade	Score	Score	CSEM	Grade	Score	Score	CSEM
4	20.0	40	2	4	24.0	50	5	4	28.0	61	2
4	20.1	40	2	4	24.1	51	5	4	28.1	61	2
4	20.2	40	2	4	24.2	52	5	4	28.2	61	2
4	20.3	40	2	4	24.3	53	4	4	28.3	61	2
4	20.4	40	2	4	24.4	54	4	4	28.4	61	2
4	20.5	40	2	4	24.5	54	3	4	28.5	61	2
4	20.6	40	2	4	24.6	55	3	4	28.6	61	2
4	20.7	41	2	4	24.7	55	3	4	28.7	61	2
4	20.8	41	2	4	24.8	55	3	4	28.8	62	2
4	20.9	41	2	4	24.9	56	3	4	28.9	62	2
4	21.0	41	2	4	25.0	56	3	4	29.0	62	2
4	21.1	41	2	4	25.1	56	2	4	29.1	62	2
4	21.2	41	2	4	25.2	57	2	4	29.2	62	2
4	21.3	41	2	4	25.3	57	2	4	29.3	62	2
4	21.4	41	2	4	25.4	57	2	4	29.4	62	2
4	21.5	42	2	4	25.5	57	2	4	29.5	63	2
4	21.6	42	2	4	25.6	57	2	4	29.6	63	2
4	21.7	42	2	4	25.7	58	2	4	29.7	63	2
4	21.8	42	2	4	25.8	58	2	4	29.8	63	2
4	21.9	42	2	4	25.9	58	2	4	29.9	63	2
4	22.0	42	2	4	26.0	58	2	4	30.0	63	2
4	22.1	43	2	4	26.1	58	2	4	30.1	63	2
4	22.2	43	2	4	26.2	58	2	4	30.2	64	2
4	22.3	43	2	4	26.3	58	2	4	30.3	64	2
4	22.4	43	2	4	26.4	59	2	4	30.4	64	2
4	22.5	43	2	4	26.5	59	2	4	30.5	64	2
4	22.6	43	2	4	26.6	59	2	4	30.6	64	2
4	22.7	44	2	4	26.7	59	2	4	30.7	65	2
4	22.8	44	2	4	26.8	59	2	4	30.8	65	2
4	22.9	44	2	4	26.9	59	2	4	30.9	65	2
4	23.0	44	3	4	27.0	59	2	4	31.0	65	3
4	23.1	45	3	4	27.1	60	2	4	31.1	66	3
4	23.2	45	3	4	27.2	60	2	4	31.2	66	3
4	23.3	45	3	4	27.3	60	2	4	31.3	66	3
4	23.4	46	3	4	27.4	60	2	4	31.4	67	3
4	23.5	46	3	4	27.5	60	2	4	31.5	67	3
4	23.6	47	4	4	27.6	60	2	4	31.6	68	4
4	23.7	47	4	4	27.7	60	2	4	31.7	68	4
4	23.8	48	5	4	27.8	60	2	4	31.8	69	5
4	23.9	49	5	4	27.9	61	2	4	31.9	70	8
								4	32.0	70	8

	Composite	Scale			Composite	Scale			Composite	Scale	
Grade	Score	Score	CSEM	Grade	Score	Score	CSEM	Grade	Score	Score	CSEM
8	8.0	1	7	8	12.0	16	2	8	16.0	27	6
8	8.1	7	7	8	12.1	16	2	8	16.1	28	5
8	8.2	8	5	8	12.2	16	2	8	16.2	29	5
8	8.3	9	4	8	12.3	17	2	8	16.3	30	4
8	8.4	10	4	8	12.4	17	2	8	16.4	30	4
8	8.5	10	3	8	12.5	17	2	8	16.5	31	3
8	8.6	11	3	8	12.6	17	2	8	16.6	31	3
8	8.7	11	3	8	12.7	17	2	8	16.7	32	3
8	8.8	11	3	8	12.8	17	2	8	16.8	32	3
8	8.9	12	2	8	12.9	17	2	8	16.9	32	3
8	9.0	12	2	8	13.0	17	2	8	17.0	33	2
8	9.1	12	2	8	13.1	17	2	8	17.1	33	2
8	9.2	12	2	8	13.2	18	2	8	17.2	33	2
8	9.3	13	2	8	13.3	18	2	8	17.3	33	2
8	9.4	13	2	8	13.4	18	2	8	17.4	34	2
8	9.5	13	2	8	13.5	18	2	8	17.5	34	2
8	9.6	13	2	8	13.6	18	2	8	17.6	34	2
8	9.7	13	2	8	13.7	18	2	8	17.7	34	2
8	9.8	14	2	8	13.8	18	2	8	17.8	34	2
8	9.9	14	2	8	13.9	18	2	8	17.9	34	2
8	10.0	14	2	8	14.0	19	2	8	18.0	35	2
8	10.1	14	2	8	14.1	19	2	8	18.1	35	2
8	10.2	14	2	8	14.2	19	2	8	18.2	35	2
8	10.3	14	2	8	14.3	19	2	8	18.3	35	2
8	10.4	14	2	8	14.4	19	2	8	18.4	35	2
8	10.5	15	2	8	14.5	19	2	8	18.5	35	2
8	10.6	15	2	8	14.6	20	2	8	18.6	36	2
8	10.7	15	2	8	14.7	20	2	8	18.7	36	2
8	10.8	15	2	8	14.8	20	2	8	18.8	36	2
8	10.9	15	2	8	14.9	20	2	8	18.9	36	2
8	11.0	15	2	8	15.0	20	2	8	19.0	36	2
8	11.1	15	2	8	15.1	21	2	8	19.1	36	2
8	11.2	15	2	8	15.2	21	3	8	19.2	36	2
8	11.3	15	2	8	15.3	21	3	8	19.3	36	2
8	11.4	16	2	8	15.4	22	3	8	19.4	37	2
8	11.5	16	2	8	15.5	22	3	8	19.5	37	2
8	11.6	16	2	8	15.6	23	3	8	19.6	37	2
8	11.7	16	2	8	15.7	23	4	8	19.7	37	2
8	11.8	16	2	8	15.8	24	5	8	19.8	37	2
8	11.9	16	2	8	15.9	25	5	8	19.9	37	2

Appendix H: Composite to Scale Score Tables - Grade 8

	Composite	Scale			Composite	Scale			Composite	Scale	
Grade	Score	Score	CSEM	Grade	Score	Score	CSEM	Grade	Score	Score	CSEM
8	20.0	37	2	8	24.0	48	6	8	28.0	61	2
8	20.1	37	2	8	24.1	50	6	8	28.1	61	2
8	20.2	37	2	8	24.2	51	5	8	28.2	61	2
8	20.3	38	2	8	24.3	52	4	8	28.3	61	2
8	20.4	38	2	8	24.4	53	4	8	28.4	61	2
8	20.5	38	2	8	24.5	53	3	8	28.5	61	2
8	20.6	38	2	8	24.6	54	3	8	28.6	62	2
8	20.7	38	2	8	24.7	54	3	8	28.7	62	2
8	20.8	38	2	8	24.8	55	3	8	28.8	62	2
8	20.9	38	2	8	24.9	55	3	8	28.9	62	2
8	21.0	38	2	8	25.0	55	3	8	29.0	62	2
8	21.1	39	2	8	25.1	55	2	8	29.1	62	2
8	21.2	39	2	8	25.2	56	2	8	29.2	62	2
8	21.3	39	2	8	25.3	56	2	8	29.3	63	2
8	21.4	39	2	8	25.4	56	2	8	29.4	63	2
8	21.5	39	2	8	25.5	56	2	8	29.5	63	2
8	21.6	39	2	8	25.6	57	2	8	29.6	63	2
8	21.7	39	2	8	25.7	57	2	8	29.7	63	2
8	21.8	40	2	8	25.8	57	2	8	29.8	63	2
8	21.9	40	2	8	25.9	57	2	8	29.9	64	2
8	22.0	40	2	8	26.0	57	2	8	30.0	64	2
8	22.1	40	2	8	26.1	58	2	8	30.1	64	2
8	22.2	40	2	8	26.2	58	2	8	30.2	64	2
8	22.3	40	2	8	26.3	58	2	8	30.3	64	2
8	22.4	40	2	8	26.4	58	2	8	30.4	65	2
8	22.5	41	2	8	26.5	58	2	8	30.5	65	2
8	22.6	41	2	8	26.6	58	2	8	30.6	65	2
8	22.7	41	2	8	26.7	59	2	8	30.7	65	2
8	22.8	41	2	8	26.8	59	2	8	30.8	66	2
8	22.9	42	2	8	26.9	59	2	8	30.9	66	2
8	23.0	42	2	8	27.0	59	2	8	31.0	66	3
8	23.1	42	3	8	27.1	59	2	8	31.1	66	3
8	23.2	42	3	8	27.2	59	2	8	31.2	67	3
8	23.3	43	3	8	27.3	60	2	8	31.3	67	3
8	23.4	43	3	8	27.4	60	2	8	31.4	67	3
8	23.5	43	3	8	27.5	60	2	8	31.5	68	3
8	23.6	44	4	8	27.6	60	2	8	31.6	68	4
8	23.7	45	4	8	27.7	60	2	8	31.7	69	4
8	23.8	45	5	8	27.8	60	2	8	31.8	70	5
8	23.9	47	6	8	27.9	60	2	8	31.9	70	7
								8	32.0	70	7

	Composite	Scale			Composite	Scale			Composite	Scale	
Grade	Score	Score	CSEM	Grade	Score	Score	CSEM	Grade	Score	Score	CSEM
11	8.0	1	8	11	12.0	17	2	11	16.0	25	3
11	8.1	5	8	11	12.1	17	2	11	16.1	26	3
11	8.2	7	5	11	12.2	17	2	11	16.2	26	3
11	8.3	8	4	11	12.3	17	2	11	16.3	27	3
11	8.4	9	4	11	12.4	17	2	11	16.4	27	3
11	8.5	9	4	11	12.5	17	2	11	16.5	28	3
11	8.6	10	3	11	12.6	17	2	11	16.6	28	3
11	8.7	10	3	11	12.7	18	2	11	16.7	28	3
11	8.8	11	3	11	12.8	18	2	11	16.8	29	3
11	8.9	11	3	11	12.9	18	2	11	16.9	29	3
11	9.0	11	3	11	13.0	18	2	11	17.0	29	2
11	9.1	11	3	11	13.1	18	2	11	17.1	29	2
11	9.2	12	2	11	13.2	18	2	11	17.2	30	2
11	9.3	12	2	11	13.3	18	2	11	17.3	30	2
11	9.4	12	2	11	13.4	19	2	11	17.4	30	2
11	9.5	12	2	11	13.5	19	2	11	17.5	30	2
11	9.6	13	2	11	13.6	19	2	11	17.6	31	2
11	9.7	13	2	11	13.7	19	2	11	17.7	31	2
11	9.8	13	2	11	13.8	19	2	11	17.8	31	2
11	9.9	13	2	11	13.9	19	2	11	17.9	31	2
11	10.0	13	2	11	14.0	20	2	11	18.0	31	2
11	10.1	14	2	11	14.1	20	2	11	18.1	31	2
11	10.2	14	2	11	14.2	20	2	11	18.2	32	2
11	10.3	14	2	11	14.3	20	2	11	18.3	32	2
11	10.4	14	2	11	14.4	20	2	11	18.4	32	2
11	10.5	14	2	11	14.5	21	2	11	18.5	32	2
11	10.6	14	2	11	14.6	21	2	11	18.6	32	2
11	10.7	15	2	11	14.7	21	2	11	18.7	32	2
11	10.8	15	2	11	14.8	21	2	11	18.8	33	2
11	10.9	15	2	11	14.9	21	2	11	18.9	33	2
11	11.0	15	2	11	15.0	22	2	11	19.0	33	2
11	11.1	15	2	11	15.1	22	3	11	19.1	33	2
11	11.2	15	2	11	15.2	22	3	11	19.2	33	2
11	11.3	16	2	11	15.3	23	3	11	19.3	33	2
11	11.4	16	2	11	15.4	23	3	11	19.4	33	2
11	11.5	16	2	11	15.5	23	3	11	19.5	34	2
11	11.6	16	2	11	15.6	24	3	11	19.6	34	2
11	11.7	16	2	11	15.7	24	3	11	19.7	34	2
11	11.8	16	2	11	15.8	25	3	11	19.8	34	2
11	11.9	16	2	11	15.9	25	3	11	19.9	34	2

Appendix I: Composite to Scale Score Tables - Grade 11

	Composite	Scale			Composite	Scale			Composite	Scale	
Grade	Score	Score	CSEM	Grade	Score	Score	CSEM	Grade	Score	Score	CSEM
11	20.0	34	2	11	24.0	46	6	11	28.0	59	2
11	20.1	34	2	11	24.1	47	5	11	28.1	60	2
11	20.2	35	2	11	24.2	48	5	11	28.2	60	2
11	20.3	35	2	11	24.3	49	4	11	28.3	60	2
11	20.4	35	2	11	24.4	50	4	11	28.4	60	2
11	20.5	35	2	11	24.5	51	4	11	28.5	61	2
11	20.6	35	2	11	24.6	51	3	11	28.6	61	2
11	20.7	35	2	11	24.7	51	3	11	28.7	61	2
11	20.8	35	2	11	24.8	52	3	11	28.8	61	2
11	20.9	36	2	11	24.9	52	3	11	28.9	61	2
11	21.0	36	2	11	25.0	53	3	11	29.0	62	2
11	21.1	36	2	11	25.1	53	3	11	29.1	62	2
11	21.2	36	2	11	25.2	53	3	11	29.2	62	2
11	21.3	36	2	11	25.3	53	3	11	29.3	62	2
11	21.4	36	2	11	25.4	54	2	11	29.4	62	2
11	21.5	36	2	11	25.5	54	2	11	29.5	63	2
11	21.6	37	2	11	25.6	54	2	11	29.6	63	2
11	21.7	37	2	11	25.7	54	2	11	29.7	63	2
11	21.8	37	2	11	25.8	55	2	11	29.8	63	2
11	21.9	37	2	11	25.9	55	2	11	29.9	64	2
11	22.0	37	2	11	26.0	55	2	11	30.0	64	2
11	22.1	37	2	11	26.1	55	2	11	30.1	64	2
11	22.2	38	2	11	26.2	55	2	11	30.2	64	2
11	22.3	38	2	11	26.3	56	2	11	30.3	64	2
11	22.4	38	2	11	26.4	56	2	11	30.4	65	2
11	22.5	38	2	11	26.5	56	2	11	30.5	65	2
11	22.6	39	2	11	26.6	56	2	11	30.6	65	2
11	22.7	39	2	11	26.7	57	2	11	30.7	65	3
11	22.8	39	2	11	26.8	57	2	11	30.8	66	3
11	22.9	39	3	11	26.9	57	2	11	30.9	66	3
11	23.0	40	3	11	27.0	57	2	11	31.0	66	3
11	23.1	40	3	11	27.1	57	2	11	31.1	67	3
11	23.2	40	3	11	27.2	58	2	11	31.2	67	3
11	23.3	40	3	11	27.3	58	2	11	31.3	67	3
11	23.4	41	3	11	27.4	58	2	11	31.4	68	3
11	23.5	41	3	11	27.5	58	2	11	31.5	68	4
11	23.6	42	4	11	27.6	59	2	11	31.6	69	4
11	23.7	43	4	11	27.7	59	2	11	31.7	70	5
11	23.8	43	5	11	27.8	59	2	11	31.8	70	5
11	23.9	45	5	11	27.9	59	2	11	31.9	70	8
								11	32.0	70	8

Appendix J: Writing Techncology Events Summary

DOCUMENT 1 WRITING TECHNOLOGY EVENTS: PRESENTED AT NEBRASKA TECHNICAL ADVISORY MEETING

2015-2016 NeSA

Writing Technology Events

Nebraska Technical Advisory Meeting

August 11, 2016

This document outlines the technical issues that occurred during the 2015-2016 NeSA Writing Assessment, which took place January 18-February 5, 2016.

SUMMARY OF ISSUES

DRC can confirm four specific situations that occurred during the NeSA Writing Assessment (NeSA-W) window related to the technology utilized in online testing. The specific detail below regarding each instance has been previously provided to the Nebraska Department of Education and is included here for reference.

- 1) eDIRECT Outage: On 01/21/16, an eDIRECT outage occurred that lasted for 1 hour and 47 minutes and affected some users. There was a general slowdown, termed a "partial outage," in the network that impacted eDIRECT and related DRC INSIGHT systems. The root cause of this issue was found and attributed to a software bug in one network device from DRC's third-party networking vendor. This bug negatively impacted the performance of the firewall and caused the general slowdown of our systems. Once DRC engaged the firewall vendor to troubleshoot this issue, they responded immediately. DRC is continuing to monitor these systems daily throughout test windows to prevent this type of occurrence from happening again.
- 2) DRC INSIGHT Outage: A DRC INSIGHT outage occurred on 01/27/16 for 32 minutes and impacted all users. DRC engineers determined that the Nebraska Writing server used to store results was low on available hard disk space. The server filled up its disk space and DRC responded to the alert that it was happening, but was not able to respond in time to stop it from occurring. DRC data management staff added significant additional storage capacity immediately and have modified the response procedure for this alert in the future.
- 3) **Invalid Characters:** DRC received reports early in the NeSA-W testing window that students were entering invalid characters in their responses. These invalid characters were attributed to non-English characters entered by students with a foreign language keyboard or keyboard shortcuts. An update to

the INSIGHT web-based test engine (WBTE) was deployed overnight on 01/27/16 to block the invalid characters.

4) Dictionary Tool and Spellcheck Unavailable: The Dictionary and Spellcheck were unavailable to students testing on 01/28/16 for 4 hours 7 minutes. This impacted all potential users. Earlier in the testing window, DRC received reports of students entering invalid characters in their responses--typically attributed to non-English characters entered with a foreign language keyboard or keyboard shortcuts. An update to the INSIGHT web-based test engine (WBTE) was deployed overnight on 01/27/16 to block the invalid characters. While the invalid character issue was resolved with the change, when the updated WBTE was deployed, the configuration that enabled the Dictionary tool to be on for all students was not set correctly.

Q & A REQUESTED BY NDE

NDE requested DRC responses to several questions related to the online issues. The Q and A below has been previously provided to the Nebraska Department of Education and is included here for reference.

1) What troubleshooting steps were used to resolve the issue(s)? DRC has established processes to troubleshoot and rapidly address technology issues when they arise. In situations where DRC is notified of an incident with online testing, and it is classified as a critical priority a DRC Quick Response Team (a.k.a. Tiger Team) is immediately called. This is a team of cross-functional Level 2, Level 3, and senior leadership resources, assembled to efficiently diagnose, troubleshoot, and resolve critical incidents. For each of the issues identified with the Nebraska Writing Assessment, the Quick Response Team was assembled and took action to identify the cause of the issue and to coordinate troubleshooting, resolution, and communication.

DRC also has on-the-ground resources in-state in Nebraska to assist with technology-related issues. Mr. Ryne Keel is based out of Lincoln. He serves as an in-state resource to help respond to incidents and support DRC and Nebraska during the issue resolution process. Mr. Keel has been a part of the team and is involved in all Nebraska-related issues.

2) If these were random occurrences, what variables were common to the users having the issues? In each of these recent issues, the commonality would have been a user seeking to connect to a DRC system or tool that was not available to them.

- 3) What notification chain was used to communicate the issues? DRC's process is that when DRC receives a fourth report from the field of an incident that appears to be the same issue the following communication steps are taken once DRC can confirm and define the issue for users:
 - a) DRC notifies NDE of the situation.
 - b) DRC posts a message to the status page.
 - c) DRC emails a notice to the district contacts.
 - d) DRC sends a follow-up email when the situation has been resolved.
 - e) DRC posts an updated message to the status page.

In the case of the INSIGHT outage, DRC initiated this process after the third report from the field.

- 4) Why did the DRC Nebraska System Status website read 'Normal Operations' when Nebraska school districts were having problems? Recent incidents were posted in 17 minutes (1/21) and 13 minutes (1/28). Once DRC discovers a report pattern that suggests a concern is present, NDE is contacted and communication is agreed to. The website message is updated at that time. On 1/27 the time between DRC's first report and resolution was 32 minutes and no Status Update message was changed. Districts were emailed, and emailed again 15 minutes later.
- 5) Can school technology coordinators be added to the DRC/NDE database for immediate notification when problems occur? DRC has a process that is established by the NDE to first notify them, and agree to a message that is both posted and then emailed to school district contacts. The DRC/NE System Status site is used to communicate widespread issues impacting online testing. District NTACs (NeSA Technology Assessment Coordinators) were added to the email distribution list for these messages on 1/28.

DEFINITIONS OF SYSTEM MESSAGING

NDE has requested definition regarding several system messages, including circumstances that may trigger their display. The specific detail below regarding each message has been previously provided to the Nebraska Department of Education and is included here for reference.

1) **Configuration Not Found Messages** (occurred 1/21 during the INSIGHT partial outage): When this message appears it means that DRC INSIGHT cannot retrieve the configuration profile associated with a device because it cannot find the Device Toolkit ORG Unit ID, which was entered incorrectly, was deleted, or was not assigned to the device.

- 2) **Device Toolkit Errors** (occurred 1/21 during the eDIRECT outage): These errors are associated with one of the following situations related to the Device Tool Kit Organizational Unit:
 - The ORG Unit was deleted after the device was assigned to it, or
 - The ORG Unit was not uploaded, or
 - The ORG Unit was entered incorrectly, or it was not set up in Chrome Management or in an MDM.
- 3) **Invalid Username or Password Messages** (occurred 1/27 during the INSIGHT partial outage): While users may receive this message when their Username or Password is outdated or entered incorrectly, it may also be displayed when the system they are seeking to enter is unavailable to them.
- INSIGHT Dictionary Tool Errors (occurred 1/28 when the tool was unavailable): The Dictionary tool and Spellcheck tool were not available on 1/28 to any student that started testing before 11:07 a.m. Any word entered into the Dictionary tool returned a "Sorry, no responses were found" message.
- 5) Testing Site Manager (TSM) "Overload" Errors: This may be a user diagnosis because it not a DRC Testing Site Manager (TSM) error message and does not match any report from the field. The TSM does have a variety of user messages, found in the Technology User Guide on pages 195-208. All TSM messaging is accompanied by detail as to how to resolve an issue. These are most frequently associated with configuration or connectivity.
- 6) **Network Error Unknown Error Code (-324) Message:** This error code displays when there is an issue with interactions over the network, often due to dropped connections. This likely happened during the firewall issues that impacted eDIRECT on 1/21/16.
- 7) Please Wait Messages after Submitting a Test: In order to protect the students' responses from network latency or bandwidth issues when submitting them, INSIGHT will wait an extended period of time before it will return an error that it can't initially send the response. A 5 minute time period is set within the system to attempt to send the response back to DRC. If that fails then there is an additional 5 minutes to send the response to the TSM if one is in use. That combination can result in 10 minutes of waiting if there is a serious network issue at that site. This could have been an additional consequence of the firewall issue that occurred on 1/21/16.

AUGUST 11 TAC MEETING - ITEM 2:

NESA-WRITING 2015-2016 ISSUES OF TECHNOLOGY EVENTS AT GRADES 8 AND 11 AND RELEASE OF SCORES

Analysis of Effects of Technology Events on Writing Scores

Three technology events occurred during the winter 2016 online writing assessment. The description of the events and the groups used for analysis are as follows:

Outage 1 (January 21, 2016):

- Group 1A Students who began testing 1/21 and completed the test before/during* the outage (before 1:40 pm)
- Group 1B Students who began testing 1/21 before/during the outage and completed the test after the outage (after 1:40 pm through the end of the test window)

*We don't know precisely when the outage began, just when it was reported (11:53 am) and when it was resolved (1:40 pm). That's why Group 1A includes students who could have finished testing "before" or "during" the outage.

Outage 2 (January 27, 2016):

No one successfully signed in or submitted tests during the outage.

• Group 2 - Students who began testing 1/27 before the outage (before 11:39 am) and completed the test after the outage (from 12:10 am through the end of the window)

Outage 3 (January 28, 2016):

The Dictionary/Spellcheck tool was not available to students for whom that accommodation was intended.

- Group 3A Students with the spellcheck accommodation who began testing the morning of 1/28 during the Dictionary/Spellcheck outage (before 11:07 am).
- Group 3B Students who began testing during the outage (before 11:07 am) but without the Spellcheck accommodation.

GRADE 8:

In the following analyses, the mean of the writing scale scores for each of the outage groups as well as the mean for students who experienced normal conditions are presented first. Then the results of linear

regression with the contrast groups designed to compare each group with the normal conditions group are displayed. A second linear regression with 2016 reading scores added in as a covariate is also presented. The purpose of this analysis is to see if any differences in mean were due to differences in abilities of students in the groups. Reading scores are not a perfect predictor of writing scores, however the correlation between reading and writing is high enough to warrant using reading scores for this purpose.

	outGroup	<pre>mean(write)</pre>	sd(write)	n()
1	None	45.69	13.65	16245
2	1A	45.78	13.55	1516
3	1B	46.42	15.41	410
4	2	47.93	13.62	892
5	3A	35.13	11.21	179
б	3B	48.17	12.48	3056

Table 1. Group Means, Standard Deviations and Number of Students

Table 2. Check for significant differences between outage groups and 'None'

lm(formula = write ~ outGroup, data = rw8)

Residuals:

Min	1Q	Median	3Q	Max
-47.166	-8.687	2.313	9.313	34.866

Coefficients:

COCLETCICIC	-				
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	45.68655	0.10596	431.158	< 2e-16	* * *
outGroup1A	0.09775	0.36269	0.270	0.788	
outGroup1B	0.73052	0.67536	1.082	0.279	
outGroup2	2.24282	0.46445	4.829	1.38e-06	* * *
outGroup3A	-10.55247	1.01500	-10.397	< 2e-16	* * *
outGroup3B	2.47903	0.26630	9.309	< 2e-16	* * *
Signif. code	es: 0 `***	′ 0.001 `**	*′ 0.01	`*′ 0.05 [`]	`.′ 0.1 ` ′ 1

Table 3. Add 2016 NeSA Reading scores as a covariate

lm(formula = write ~ outGroup + read, data = rw8) Residuals: 1Q Median 3Q Min Max -51.897 -7.008 -0.177 7.028 47.876 Coefficients: Estimate Std. Error t value Pr(>|t|) Intercept) 21.104415 0.244691 86.249 < 2e-16 *** outGroup1A -0.646916 0.292465 -2.212 0.026981 * outGroup1B 0.741703 0.544254 1.363 0.172963 outGroup2 1.639730 0.374894 4.374 1.23e-05 *** outGroup3A -3.273157 0.823349 -3.975 7.05e-05 *** outGroup3B 0.810396 0.215077 3.768 0.000165 *** read 0.209336 0.001941 107.872 < 2e-16 *** ___ Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Spellcheck Accommodation sample:

Table 4. Group Means, Standard Deviations and Number of Students

	group3A	<pre>mean(write)</pre>	sd(write)	n()
1	No	32.25	12.70	776
2	Yes	35.13	11.21	179

Table 5. Check for significant differences between outage groups and 'None'

lm(formula = write ~ group3A, data = s8)
Residuals:
 Min 1Q Median 3Q Max
-35.134 -5.246 1.754 7.754 37.754
Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 32.2461 0.4463 72.249 < 2e-16 ***
group3AYes 2.8879 1.0309 2.801 0.00519 **
--Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1</pre>

Table 6. Add 2016 NeSA Reading scores as a covariate

lm(formula = write ~ group3A + read, data = s8)

Residuals: Min 1Q Median 3Q Max -48.002 -6.759 0.018 7.249 35.511

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 18.50762 1.11345 16.622 <2e-16 *** group3AYes 1.72888 0.95158 1.817 0.0696 . read 0.18030 0.01356 13.292 <2e-16 *** ---Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

FINDINGS:

We can see from the results displayed in Tables 1 and 2 that the differences in means for the first outage (both 1A and 1B) are not significantly different from the comparison group (normal conditions). The means for groups 2 and 3B are higher than the comparison group and the difference is statistically significant. This is an expected finding from group 3B since that group is defined by excluding students who need an accommodation. The finding that the students who experienced an outage (group 2) scored higher than the comparison group is unexpected. However, we will not analyze that finding any further since our interest is in scores that may be lower due to the technology events.

The students who experienced the outage in the spellcheck accommodation (Group 3A) had the greatest difference when compared to all students who were not testing during that outage. This difference is also observed when we control for the reading scores (Table 3). However, since students in this group have academic disabilities and therefore have the need for the spellcheck accommodation, it is appropriate to use similar students who also need and were provided with that accommodation as a comparison group. Tables 4, 5 and 6 show that analysis.

When we compare the impact of the spellcheck outage on students who needed that accommodation compared with similar students who were provided the accommodation, we find that the students who experienced the outage scored higher (Table 4). This difference is statistically significant (Table 5). However, when the reading scores are added, the statistical significance is reduced to < 10% (Table 6).

Note that in Table 3, group 1A (students who completed testing before or during the 1/21 outage) showed lower scores controlled for reading than the comparison group and that this result is significant at the .05 level. However, since the uncontrolled mean score for this group was slightly higher than the comparison group (Table 2), it is reasonable to assume that the observed difference is due to unaccounted for random variance.

Grade 11:

The same analyses conducted for grade 8 were repeated for grade 11.

 Table 7. Group Means, Standard Deviations and Number of Students

	outGroup	<pre>mean(write)</pre>	sd(write)	n()
1	None	45.29	14.51	16709
2	1A	45.68	14.01	1858
3	1B	43.64	13.71	428
4	2	48.35	12.86	595
5	3A	34.13	11.69	77
б	3B	46.69	12.73	2019

Table 8. Check for significant differences between outage groups and 'None'

lm(formula = write ~ outGroup, data = rw11)

Residuals:

Min	1Q	Median	3Q	Max
-48.353	-8.287	0.713	10.307	35.870

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)				
(Intercept)	45.2874	0.1102	410.913	< 2e-16	* * *			
outGroup1A	0.3902	0.3484	1.120	0.2627				
outGroup1B	-1.6472	0.6974	-2.362	0.0182	*			
outGroup2	3.0656	0.5943	5.158	2.52e-07	* * *			
outGroup3A	-11.1575	1.6273	-6.857	7.24e-12	* * *			
outGroup3B	1.4055	0.3357	4.187	2.83e-05	* * *			
Signif. cod	es: 0 `**	**′ 0.001 `*	**′ 0.01	`*′ 0.05	`.′	0.1	`	'

Table 9. Add 2016 NeSA Reading scores as a covariate

lm(formula = write ~ outGroup + read, data = rw11)

Residuals:

Min 1Q Median 3Q Max -60.906 -7.043 -0.282 7.406 45.654

Coefficients:								
	Estimate	Std. Error	t value	Pr(> t)				
(Intercept)	24.346027	0.219219	111.058	< 2e-16	* * *			
outGroup1A	-0.228365	0.279926	-0.816	0.414622				
outGroup1B	-0.851146	0.558715	-1.523	0.127674				
outGroup2	1.321365	0.474557	2.784	0.005367	* *			
outGroup3A	-4.307926	1.299113	-3.316	0.000915	* * *			
outGroup3B	0.559115	0.268964	2.079	0.037650	*			
read	0.192422	0.001819	105.807	< 2e-16	* * *			
Signif. code	es: 0 `**;	*′ 0.001 `*;	*′ 0.01	*′ 0.05	`.′ 0.1 ` ́			

Spellcheck Accommodation sample:

 Table 10.
 Group Means, Standard Deviations and Number of Students

	group3A	mean(write)	sd(write)	n()
1	No	32.11	13.16	491
2	Yes	34.36	11.60	76

Table 11. Check for significant differences

lm(formula = write ~ group3A, data = s11) Residuals: 3Q Min 10 Median Max -33.355 -7.112 0.888 7.888 37.888 Coefficients: Estimate Std. Error t value Pr(>|t|) 0.5852 54.877 <2e-16 *** (Intercept) 32.1120 group3AYes 2.2432 1.5983 1.404 0.161 ___ Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Table 12. Add 2016 NeSA Reading as covariate

lm(formula = write ~ group3A + read, data = s11)

Residuals:

Min 1Q Median 3Q Max -37.724 -6.837 0.233 7.536 40.235

Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 19.58461 1.19626 16.37 <2e-16 *** group3AYes 1.23202 1.43246 0.86 0.39 read 0.18509 0.01572 11.78 <2e-16 *** ---Signif. codes: 0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

FINDINGS

The Grade 11 results are very similar to those from grade 8. The only group with a lower mean compared to students unaffected by the outages was the students who should have been provided a spellcheck accommodation (group 3A). However, when that group is compared to similar students who were provided that accommodation, we again see a higher mean score that was not statistically significant when controlled by the reading scores.

We again find a slightly lower mean score for students who experienced an outage on Jan 21 (group 1B), but the difference when the reading scores were added was not significant.

CONCLUSION

Although the technology events students experienced this winter were certainly disruptive to the assessment environment, the results indicate that student writing scores were not negatively impacted. It is possible that the students might have scored higher if testing had proceeded without incident, but that possibility cannot be evaluated with the data we have available.

DOCUMENT 2 NESA-WRITING VERIFICATIONS: SPELLCHECKER PAPERS

During NeSA-Writing Verifications, 260 student essays were reviewed to investigate the effect of interrupted access to the spellchecker feature on the NeSA-Writing test. All essays reviewed were written by students with IEP plans that stated they should receive access to the online spellchecker accommodation on NeSA-Writing. Each essay was reviewed during the verification process to determine if the score for Sentence Fluency & Conventions would be considered for a higher score if the student had had access to the spellchecker feature for the duration of the test. The verification committee reviewed the essay responses in pairs and flagged essays that required whole group review.

In 8th grade 183 essays were reviewed, of which four were flagged for further review and whole group discussion. After committee discussion, it was determined that access to spellcheck would not likely have corrected the majority of the spelling errors on the four essays, as the errors were in usage--for example, the use of "dune" and "don" for the word "done." The committee also determined that the overall sentence fluency on all 4 papers fell within the score point originally assigned. No score changes were made to any reviewed essays at 8th grade.

In grade 11, 77 essays were reviewed, and none was flagged for further review. No score changes were made.

Overall, the Nebraska educators on the Verifications Committee determined that the scores assigned to all 260 essays accurately reflected the students' demonstration of Sentence Fluency and Conventions, even when spelling errors were removed from consideration.

DOCUMENT 3

<u>BACKGROUND</u>: When scores were released to school districts for NeSA-Writing in April 2016, Nebraska school districts questioned the validity of NeSA-Writing scoring for grades 4, 8, and 11 for 2015-2016. Districts claimed the scoring was inconsistent, and sent NDE student essays that they believe showed less skillfully written essays receiving higher scores than essays written at a higher level. NDE advised districts to submit all essays they considered miss-scored for NeSA-Writing verification.

In addition, adding to the perception that essays were not scored correctly, district personnel contacted NDE to report that scores across domains were the same for many essays. The districts claimed that too many essays received 4, 4, 4, 4, or 6, 6, 6, 6,--and stated they were of the opinion that DRC scorers did not give enough consideration to student performance on individual domains of the rubric, but rather, essentially, assigned holistic scores by giving the same scores across all domains, resulting in lowered overall scores because students were not rewarded for doing well in individual domains.

Finally, adding to the perception of misscoring were the statewide results, as all three grades-- 4, 8, and 11 decreased in the percent proficient from 2015-2016.

NeSA-Writing	Grade 4	Grade 8	Grade 11
2014-2015	70%	71%	76%
2015-2016	68%	68%	70%

Table D.3.1 Percent Proficient

<u>NESA-WRITING VERIFICATION PROCESS</u>: NeSA-Writing scores for individual students, schools, and districts are provided to Nebraska districts through eDIRECT in April of each year. School staff members are able to review individual student scores, and if they would like to have an essay reviewed by committee of Nebraska reviewers, submit to NDE a justification of the score change being requested. Only essays that receive scores that place them in the Not Met category can be submitted for verification.

NDE compiles all verification requests and in June of each year convenes a committee of Nebraska educators who are leaders in the state in the area of writing and bring much experience and expertise in writing instruction and scoring. The Nebraska Department of Education Writing Director convenes the committee and oversees all processes. Verification committee members are trained with the same training packets used at Data Recognition Corporation during the DRC NeSA-Writing scoring process. Then, working in pairs, members of the committee review each essay and the justification provided by the school district to determine if the requested score change is appropriate. A data entry person works on site and as scores changes are approved or denied, enters all data changes. Individual districts receive a report on all submitted verifications, indicating whether each essay had a score change or not.

Each NeSA-Writing domain is worth 4 points; therefore, since each essay is scored by two DRC graders, the raw score range is 8-16. An essay that received a 1 from each grader in all four domains would earn an 8. An essay that received a 4 from each grader in all four domains would receive a 16. The raw score is converted to a scale score based on the Raw to Scale Score Conversion Charts.

Almost every score that is submitted for a review is a 2/3 split in the domain for which verification is being requested—and in most cases, the writer of the verification defense is asking for the essay to be a 3/3 in the domain, which would result in a passing score for the student. A 2/3 split is an acceptable adjacent score on a scoring rubric of 1-4. At grades 4, 8, and 11, a student who earns all 4s from both graders will pass, a student who earns all 2s will not pass. How many 2s a student can earn and still pass varies by grade level, with grade 11 cut score allowing the least number of 2s and grade 4 allowing the most 2s.

It is difficult to ascertain exactly how many 2s a student may have as the cut scores vary by grade, and the domains are weighted. Since Ideas and Contest is worth 35% of the final score, a 2 in Ideas and Content has a larger effect on a student's final score. NDE has posted a Writing Raw Score Calculator on its website, which schools personnel use extensively to determine how the grader ratings result in final scores for students.

Writing Raw Score	Writing Raw Score	Writing Raw Score
Calculatorxlsx version	Calculatorxlsx version	Calculatorxls version

Results of the 2014-2015 NeSA-Writing Verification Process

Grade	4	8	11
Number of essays submitted	1011	713	313
Number of scores changed	313	253	129
Percent of verification submissions changed	.31	.35	.41
Total NeSA-W tests per grade	22,636	21,919	21,223
Percent of score changes of total essays per grade	.014	.012	.006

2015-2016 NeSA-Writing Verification Results

Grade	4	8	11
Number of essays submitted	1,205	994	622
Number of scores changed	418	393	221
Percent of verification submissions changed	.35	.40	.36
Total NeSA-W tests per grade	23,088	22,261	21,516
Percent of score changes of total essays per grade	.018	.018	.010

Document 4

HOLISTIC SCORING VERSUS ANALYTIC SCORING

NDE is exploring the possibility of analytically scoring the text-dependent analytics to be included on the NeSA-English Language Arts test at grades 5-8. Analytic scoring that results in highly correlated scores across the domains are limited in giving guidance to Nebraska educators for improving student learning. Results of the Nebraska holistic scoring and analytic scoring of samples of essays may inform the usefulness of analytic scoring.

<u>Holistic Scoring</u>: Holistic rangefinders were established through a rangefinding process, conducted by NDE and DRC with Nebraska educators serving as readers to establish the range. After the rangefinders were established, a sample of 1500 essays for each prompt at grades 5-8 and 11 was scored holistically, with each essay receiving a single reading, by Data Recognition Corporation scorers at the company's Minnesota scoring center, using the rangefinders established through the process with Nebraska educators.

<u>Analytic Scoring:</u> NDE and Education Service Unit 3 partnered to lead an analytic rangefinding process to establish analytic ranges for all three domains of the rubric with the Nebraska educators serving as readers to establish the ranges. The analytic rangefinder readers were not the same group of educators as the holistic rangefinder readers.

Two prompts were selected from the five scored holistically at each grade, and the same 1500 essays for each prompt were scored analytically by the same group of Nebraska educators, that established the analytic rangefinders.

The same rubric was used for both holistic and analytic scoring.