

Technical Advisory Committee
Nebraska Department of Education
April 24, 2013

Nebraska Department of Education, Lincoln, NE 3:00-4:00 pm (CST)
V-TEL Videoconference Meeting

Video Conference Locations:

- Nebraska Dept. of Education, 301 Centennial Mall South, Lincoln NE, 6th floor, V-tel room (NDE Statewide Assessment staff and Harwood)
- 45 Columbus Ave., New York City, NY (Camara)
- 500 ACT Drive, Iowa City, IA (Sawyer)
- 31 Mount Vernon St., Dover, NH (Gong)
- 6949 So. 110th ST., Omaha NE (Poole)
- 1126 NW 29th St., Corvallis OR (Auty)

3:00 PM Welcome and Introductions- Valorie

In attendance: Chair- Brian Gong, Richard Sawyer, Bill Auty, Wayne Camarra, Linda Poole, Valorie Foy, John Moon, Edward Foy, Frank Harwood, Jeremy Heneger

Guests: Mark Feldhausen, Tami Williams—Millard Public Schools

3:05 PM Approve Minutes – Brian

Minutes Approved

Linda Poole moved to approve minutes from the TAC meeting held November 7, 2012

Frank Harwood seconded

Brian Gong suggested a correction. In the following statement from the minutes, Brian suggested *larger schools* should be replaced by *smaller schools*. Richard affirmed that the statement should read *smaller schools*.-“Richard Sawyer commented on the interpretation of improvement scores: what is the extent by which improved scores can be an indication that the improvement is not by chance. Most observed changes are not statistically significant (such as fluctuation in ACT scores in larger schools).”
Approved

3:05-4:00 PM Update: Nebraska State Accountability [NeSA]-Writing 2012-2013 Valorie

The 2012-2013 NeSA- Writing testing window took place January 21 through February 8, 2013. The test engine on the writing test caused some formatting errors for students in grades eight and eleven. NDE did receive reports of considerable frustration on the part of some students due to the formatting. The on-line issue was not resolved over the course of the three-week window. For the evaluation of the essays, Data Recognition Corporation [DRC] created a training packet and trained the DRC graders for a “fifth domain,” in which the essay evaluators indicated for each essay if a formatting error appeared. The “fifth domain” did not contribute to the essay’s score but merely serve as identification of essays that had formatting errors. Graders were directed to evaluate the essays without penalizing students for any formatting errors, and if any grader had a problem in evaluating a particular essay due to formatting errors to pass the essay to a team leader. “Presentation” is not a domain by which essays have ever been evaluated on NeSA-Writing. The percentage of essays that received a flag from

both graders indicating formatting errors is 45.3% at eighth grade and 42.6% at eleventh grade. The following decisions have been made by NDE:

- To release the NeSA-Writing scores.
- To indicate on the score release reports and on the State of the Schools Report that formatting issues occurred in the on-line engine at grades 8 and 11.
- To **not** include the NeSA-W status scores for grades 8 and 11 in Nebraska Performance Accountability System [NePAS].
- To **not** include the NeSA-W scores in improvement for grades 8 and 11 in NePAS for the 2013 State of the Schools Report.

Documents Attached

- DRC Training Packet
- NeSA Writing 2013 – Formatting Analysis by Bill Auty
- Sample emails from school districts

Question: Have any TAC members had a similar experience in testing that could inform this situation, where students experienced frustration due to the on-line engine not performing correctly? What suggestions does the TAC have for informing districts and the public about the situation with the writing scores? What recommendations does TAC have in regard to using the 2012-2013 scores in comparison to the 2013-2014 scores to measure improvement? Would the improvement in school district scores for grades 8 and 11 in 2012 and for grades 8 and 11 in 2014 be a acceptable measure of improvement? Grade 4 scores for improvement would compare 2013 and 2014. Have TAC members had experience with procedure for payment for the on-line services to a company where such a situation arose that could inform NDE's decision on payment?

Valorie Foy summarized the situation that arose during the NeSA-Writing 2013 testing at grades 8 and 11, which took place January 21-February 8, 2013

Districts reported some minor issues arising during the field testing window in November-December 2012. NDE participated in discussions concerning these issues with Data Recognition Corporation [DRC] and its subcontractor Computerized Assessment Learning [CAL], the company that provided the writing test engine. It was recommended by DRC/CAL that some suggestions be made to districts, such as not using formatting tools—for example, font size, indentation, and centering. NDE did not agree to the suggestions as they would have changed the parameters of the test just before the testing window opened, in addition to possibly making teachers and students uncomfortable. NDE expected DRC and CAL to have an engine that operated at least as well as it had in 2011-2012, a testing year in which students experienced some minor formatting issues, but nothing that raised concern for students, teachers, or school districts.

However, as the NeSA-Writing test window opened at the end of January, issues were reported concerning the test engine—and as the testing window progressed, the issues became more widespread. Students experienced irregularities such as lines that extended past the right margin where a student could not read the line and formatting selections that became “stuck,”

such as centering that would not turn off so the student's entire essay centered. Other issues were less intrusive, such as a word wrapping oddly or one word appearing in a different font than the rest of the essay—and students were unable to correct the irregularities when they tried to. The formatting issues occurred in what appeared to be an erratic manner as some districts experienced almost no issues and others had entire classrooms that experienced some issues with formatting. In addition, NDE was informed by DRC/CAL that formatting errors would not save and would not show up in student essays, but eventually this information was determined to be erroneous. CAL provided a fix to the software engine part way through window, but NDE noted a number of issues that resulted in a decision to not provide the update to school districts:

- The update was not guaranteed as a complete fix
- Districts would have to push out the update onto student machines
- Having some districts experiencing issues and others testing after a fix had occurred appeared inequitable.

NDE did not opt to ask CAL to push the fix out.

In addition staff at Papillion LaVista had notified DRC about the problems and received a guide to fix the problems; the suggestions did not work.

The testing window closed with a number of districts having contacted NDE and DRC concerning the formatting issues.

Wayne Camarra indicated that software changes in the middle of an operational test are not advisable, that NDE should want to know what changes were introduced that caused the issues, and the company should provide their plans for quality assurance in the future.

After the window closed DRC proposed a method of identifying essays that showed evidence that a student may have experienced formatting issues. DRC trained graders in a “fifth domain,” asking raters to identify essays in which they could see formatting irregularities. This was the most accurate method available as the engine itself did not identify issues that occurred in formatting. NDE understood that there may have been over-identification of formatting issues as students may have created some formatting errors themselves; however, this method was the only measure available.

Bill Auty provided information on the research he conducted using DRC's method of flagging essays. Students who were flagged tended to have higher scores.

Richard Sawyer had consulted with writing specialist in his organization and noted: The students who were flagged had higher scores than the students who did not. The students who were better writers who have higher scores as a rule write longer essays and those who write longer essays are more likely to have encountered formatting so there is a kind of correlation effect.

Valorie affirmed the premise: Anecdotally, the districts reported that students who edited a lot experienced more errors. Those who did not edit were less likely to have formatting errors.

Brian Gong noted an alternate hypothesis might be that scorers noticed the irregularities as they flagged them and were affected to score more positively.

Linda Poole asked if the formatting issues would have impacted a special population more. NDE had requested further information from DRC and did not find that any particular group was disadvantaged.

Bill indicated that to use the 7th grade reading was the best proxy for determining students' ability in 8th grade writing. Even if one equalized by 7th grade reading scores, students whose essays are flagged still seemed to do better.

Valorie indicated that NeSA-Writing scores at grades 8 and 11 were to be included on the State of the Schools Report but not included on NePAS. The performance level (percent proficient) had improved for both grades, although less at grade 8 than grade 11.

Brian noted that whatever problems occurred did depress scores, not raise them, so it was wise to not include NeSA-W at grades 8 and 11 on the Nebraska Performance Accountability System [NePAS].

John Moon noted that NDE could not assume all scores are depressed. We assume that students who did not have errors got exactly the score they would have.

NDE might use 2013 NeSA-W scores to measure improvement in 2014. If the scores were depressed in 2013, the results would give an overestimate, so would not be a disadvantage for any school district.

Frank Harwood noted that inflated improvement scores in 2014 could be a disadvantage to districts in 2015.

Frank asked if scores could have been raised by scorers who evaluated the essays after first identifying whether the student experienced some formatting issues, thus possibly inflating the scores of flagged essays.

Linda noted that students had a hard time due to frustration and students being upset can affect more than formatting issues. If NeSA-W was not to be included on NePAS, community members may not understand so NDE had an obligation to educate districts and the public.

Wayne indicated that if scores are to be included on State of the Schools Report, even if not on NePAS, NDE needed to include a public explanation of the situation.

Also noted by Wayne was the responsibility of the vendor and the need for quality assurance, even with last minute changes. DRC was responsible for its subcontractor, and it was difficult for NDE to determine who was responsible.

Brian summarized the agreed upon actions of NDE:

- To release NeSA-Writing scores at grades 8 and 11
- To flag or asterisk scores and explain in a very short sentence what happened, avoiding the word “error.”
- To release scale scores with notification also.
- To notify individual students of scores through the Individual Student Reports as has been done for the last four years-- with formatting issues noted.

A discussion ensued about releasing the information to districts about the number of their students who were flagged as having formatting errors. A number of issues were indicated. Transparency is best. However, districts that had no flagged students might claim their scores were valid and want them included in NePAS. In addition, NDE did not have an accurate count of the number of students who experienced formatting issues, but only the number of essays that had formatting errors, errors that may have occurred through human error. NDE would have the responsibility to conduct further research if planning to release information to each district concerning number of students who experienced errors, research almost impossible to conduct.

The committee determined to discuss improvement measurement of NeSA-W at Grades 8 and 11 at next meeting.

The committee discussed repercussions for payment.

Wayne noted that NDE cannot use scores for any student for any school for any district –the portion of work that is unuseable is all of it.

Wayne indicated that NDE needed assurances that the engine was going to work well next year. However, NDE indicated that the contract had changed due to the release of an RFP in October 2012. DRC would be delivering all aspects of NeSA testing in Nebraska starting in 2013-2014, including the writing test. NDE asked for demonstrations during the RFP process and the demonstrations led to our decision.