

Technical Advisory Committee
Nebraska Department of Education
December 17, 2015

Approve Minutes of June 30, 2015-

Two changes to the minutes

ELA Assessment

NeSA-ELA Test Designs-

- Progressive test design
- Three parts to the scoring- what support will be provided to local scorers?
- ELA workshops, TDA workshops, practice TDA to select schools and scored a selection. Practice scoring was completed and feedback was given. Will conduct more workshops fall 2016
- To inform workshops, use student think alouds? Asking students what they are doing at specific steps--like a cognitive lab
- Need to engage and have teachers teach TDA and understand it. Need to know features to be able to teach--seems to be a good amount of time allotted for teacher preparation

NeSA-ELA Tables of Specifications

- Discussion of questions posed to NE educators
- The work was completed by experienced item writers, who reviewed new item types and were asked if any additional indicators should be tested and any indicators that shouldn't be tested. Feedback from teachers indicated that item types were more effective than multiple choice.
- Reporting- reports on each item, not just top 3 and bottom 3. Need to check if enough item types to do that type of analysis and also reliability of the measure psychometrically. This type of reporting could show gaps in curriculum
- Rather than have NDE make the judgment, might provide raw data to districts and let them analyze. Could have item type flag, and let the districts make the decision.
- Item type and why the item type was chosen, can be difficult for TDA when the skill set is being measured, not the actual item type.

NeSA- ELA Reporting Categories

- Discussion of five reporting categories- recommendation of extended TOS to see how many items are being reported; measurement consideration, looking for overlap.
- Rationale for breaking categories out as they are --gives more information, more challenging for students, different genre types. Grade 11- the construction of the

categories is weighted for informational, literary and vocabulary, and it changes depending on the grade range. College and career ready has a focus more on informational reading

- Might give a range on percentages or NDE might lose range of freedom

Weight of Text-Dependent Analysis

- Indications from samples in pre pilot: students are taking the task seriously, well developed responses to the TDA. Students producing a sizeable amount of text in their responses. NDE discussed limiting character amount; working with LPS to find samples using fewer words.
- Younger students will need to learn how to pre plan how much they will write, also ELL and special needs. Give consideration to weight on one single item type that will be challenging. Keyboarding is a concern. Discussion of visible character count possibly affecting the amount of students' responses. Discussion of effect of too much weight on any item.
- The 8 point weighted amounts to 4 pts reading, 4 pts writing
- Need to give consideration to weighting: comprehension and writing. Discussion of ways to weight the TDA and effect on cut scores. Reading and writing are not independent in this task.

Text-Dependent Analysis Scoring

Holistic Scoring of TDA/Analytic Scoring of TDA

- Discussion of rubric. Analytic is more helpful from instructional standpoint
- DRC indicated its scorers see a good distribution in other states, not skewed; readers are not instructed to weight any one thing more than the other. Spelling not on rubric.
- Analytic scoring captures profile- more helpful information
- The report should include profile distributions.
- A limitation of analytic may be due to combinations being the same
- Papers could be both holistically and analytically scored and compared; two way analysis may be beneficial; will have operational scores in summer 2016
- Discussion of approach used by DRC scorers and effect on costs. DRC indicates that time for analytic scoring is a factor in cost.

TDA pilot at 3rd and 4th grades- Research

- Purpose of pilot is to find about TDA and also to see if 3rd and 4th graders can work online productively; cognitive lab suggested.
- Discussion of non scoreable codes
- There is a factor of adult influence thinking that students cannot do TDA online. The study would show NE students, not nationally. Many states do 3rd grade online and

students are not being compared to handwritten test. Middle school and high school being comparable is not a consideration.

- 1) interested in differences in grades, scoring is not an issue if scorers are trained
- 2) modes- could the students do the TDA on paper as well as online? Could check literacy on other test scores.
- Students need to be familiar with the devices they are using for testing—should be typically used in classrooms

NeSA-ELA- -Alignment Study

- DRC is familiar with different alignment models. Norman Webb model was used when discussing with NE. Easier to respond to peer review guidance; there is also a tool that can be used with educators and results can be tallied very quickly.
- Alignment should be conducted when NDE can still make some changes to final operational test.
- Should encourage modification of Webb; in terms of balance, representation and range, can sometimes look askew. Strongly recommends 3-4 months ahead of final forms.
- Could check cognitive complexity very early

Equating

- Can do standard setting once item types become operational
- Discussion of scale in 2017, but the determination has not been made yet
- Equating helps with transition of scores. Consider the overlap of content; have a reasonable timeline

ELL Accommodations

- Discussion of the intent of adding “standard English”; goal is college and career ready
- The construct is English Language Arts- eliminating accommodations makes it more even
- Discussion of ELPA and ELA being required for ELL students.
- The process is right- find the construct, find threats to validity, inform variations of administration policy (accommodations)
- Construct being done by NE will challenge other states. Other states don’t require translations; require that reports be turned in in English.
- Consider why the policy is different for math and science, than ELA
- Develop rationale about why the construct is the way it is.

NeSA-Math Transition

- Consider indicators per strand when going to the lowest level-

- Could go deeper into the indicators, and sample over time such as a 3 year period so all indicators are hit. Sampling strategy; sampling is not usually public, but any sample within a stand should be decent sampling, over time.
- Not every indicator can be assessed each year. A limitation is that reporting strands may affect assumptions made at the district level which may affect curriculum. The assumptions may not be accurate.
- C4L can be used to determine how well students are doing on an indicator, but more is needed. Teacher-based reporting information would be helpful
- Could use residual analysis as residual analysis brings in more information, a different type of information; raw percentage correct is easy for people to misinterpret
- Clarification of TOS and blueprint
- Developing adequate number of items
- Increasing field test forms is feasible--either need more than 5 forms, or more usable items; going to need more items up front
- Could specify the item types, give examples and run by educators and look for items that naturally assess what you are looking for
- Implementation is very important- tools, placement, etc.

AQuESTT Accountability Sections

Raw classification-

- The impact data were changed by state board request-
- Discussion of how schools might plan school improvement, based on knowing raw numbers in the calculations

Discussion of possible changes

- Possible political effect
- Discussion of school/district size steps taken to assure system is fair and equitable across the state. Work of the Task Force. Possibility of tiering, based on size of school.
- District score is accumulation of individual scores, not school scores.
- Discussion of reports
- For a school to earn a point, the school needs quite a bit of improvement
- Consider use of norm reference adjustment, instead of error adjusted approach

Evidence-Based Analysis

- Self- reporting has a lot of impact- discussion of a plan to show how rating actually shows up in a school building. Discussion of reliability and validity of the EBA due to self-report.
- Other sources of data might be recent graduates, public input, business owners in the community. Other sources provide depth.

- Being present and present in quality is different and depending on who answered, perceptions could be very different in same school.
- Could do longitudinal study with benchmarks of public perceptions around educational
- 1) quality of the measure- write a technical report, documentation around decisions and definitions 2) Whether the answers are context dependent 3) multiple measure should do the same thing, be complimentary
- Usually in accountability, should not have small changes trigger large changes
- The way the data were presented on AQuESTT form was discussed. EBA was 2/3 of the form —and made EBA look most significant in final classification instead of the Raw Classification which had the more widespread effect.

Discussion of perception of reliability of EBA.

- Might be opportunity to acknowledge through accreditation processes; every district does have a visit every 5 years, in addition to random visits- and the processes overlap
- Could combine in an index; rule- minimum of this for this to happen, set of conditions; profile or tree decision; if they're different things it doesn't make sense to combine or put them together
- Scores may be higher next year; a rubric would be a good guide
- Consider if the data reflect a true situation, and if there is an incentive to make sure it is accurate.
- Might look at industry for a model. Consider workplace engagement, employee selection, employee satisfaction, customer satisfaction
- Some states wish they had a larger number of indicators