Oklahoma Technical Advisory Committee (OK TAC)

Oklahoma Core Curriculum Grades 3-8 & End-of-Instruction Tests

CONFIDENTIAL DRAFT

October 14, 2015

Re: TAC Feedback on Review of Draft State Content Standards

To: Sonya Fitzpatrick

From: John Olson, TAC Chair

Per your request, the Oklahoma Technical Advisory Committee (OK TAC) reviewed the Draft Version 3 of the Oklahoma Academic Standards for ELA and Mathematics, primarily from the perspective of whether the standards would be assessable in future state tests and if the strands and benchmarks can be measured by the state summative assessment. Four of the five TAC members were able to submit their comments on this. Following is the feedback from the TAC.

Overall Comments on Oklahoma Academic Standards for ELA and Mathematics

In my review, I evaluated the draft standards on issues like practicality to do, amount of time required, and costs, as well as psychometric considerations. Several things stand out regarding these new standards and how to measure some of them. For ELA, the Listening and Speaking standards will be very difficult, if not impossible, to measure in a standardized test. This can be done in a large scale assessment, such as what is done for ELP tests such as WIDA, but it is not simple to do nor inexpensive. Thus, it may be difficult to implement in the OK state summative assessment.

Some of the standards for Pre-K to Grade 2 seem to be rather challenging for young students and will be difficult to assess in a group setting. These may need to be tested in a 1:1 setting. This will require a lot more time for testing each student individually. In addition, some of the standards for Grades 3-6 ask for collaborative work and group discussions – again, this will be difficult to measure precisely and assess. Grades 7-9 have students making presentations; this also will be a challenge to do in a typical state assessment

The standards are forward thinking in their use of technology, which is good. However, the use of computers for testing of young children can be problematic if the kids are not familiar with them. The state wants to be sure that it is assessing math skills, for example, and not computer literacy skills. If young students have to use them to type their responses, this can be challenging too. It may be difficult to do with students before grade 3 (and even 3rd graders may have problems with this).

The standard for Independent Reading/Writing may also be hard to measure, since it requires this to be done over extended periods of time. I am not sure how this can be assessed in the usual format and time constraints of a typical state assessment.

Note that some of these standards may be better to assess via use of a formative/classroom-based or interim/benchmark assessment and not a summative assessment.

Concerning item types, many of the standards can be assessed using multiple choice items, but some cannot. The state will likely need to include short and extended constructed response items in its assessment. You may also need to use performance tasks for some things. The use of these types of items will require more time, reliable human scoring systems to be set up, and will also drive up the costs of the assessments.

For the Math Standards, some of the same issues apply here. Grades Pre-K to 2 include many challenging topics and require the kids to explain their answers, create Venn diagrams, problem solve and show their work, use technology, and write responses. This may be very difficult for them to do and for a statewide assessment to test. Even at grades 3-6, there will be challenges with these expectations. Some of the standards require students to perform, which MC items may have a hard time measuring. The use of CR and PT items may be needed here too, as well as the use of technologically-enhanced items, which will increase costs and time for administering the tests. Other grades require students to construct various things, display data, make graphs, model solutions, etc. Again, these standards will need to be carefully assessed in appropriate and doable ways.

One other area that needs to be considered is assessing Students with Disabilities (SWDs) and English Language Learners (ELLs) on the new standards. On some of the standards this may be hard to do with students from these populations. I suggest that the state have experts on SWDs and ELLs review the standards on how they can be measured for these students with special needs and if they will be able to access, or not access, the expectations. Also, the use of accommodations and other tools that can help with accessibility will need to be considered in this review.

More specific comments on the standards are provided below.

John Olson, TAC Chair

President, Olson Educational Measurement and Assessment Services

Math

MAPs are good for instructional focus but difficult to assess. For example, anything about disposition is difficult to measure. Fluency tends to be measured in response time, but speeded tests are not typically used in K-12 assessments. I'm not sure what a "flexible" understanding is -- maybe we could assess it by measuring the same construct in multiple ways?

Benchmarks that include phrases such as "assess the reasonableness of result" (3.N.2.4; 4.N.1.5) beg for a performance task with a written response. Great items, but expensive to develop and score.

In "**4.GM.3.3** Develop and use formulas to determine the area of rectangles. Justify why length and width are multiplied to find the area of a rectangle by breaking the rectangle into one unit by one unit squares and viewing these as grouped into rows and columns."

And others like it (e.g., ELAPA.GM.1.1), I'm not sure how I would ask a student to "justify why" on an assessment. Great for classroom assignments and discussion, but might be difficult to measure.

In general, it is difficult to know how a student solved a problem in a traditional assessment, unless they are asked to explain their answer. A lot of the benchmarks indicate a student is supposed to use a specific approach to solve a problem. Math items can be set up to determine whether a student can solve a problem using a particular approach, but it is more difficult to determine if they will choose to use that approach if given freedom to take any approach.

Again: **7.D.1.1** Design simple experiments, collect data and calculate measures of central tendency (mean, median, and mode) and spread (range). Use these quantities to draw conclusions about the data collected and make predictions. Begs for a performance task -- great tool, but expensive.

ELA

Speaking and Listening will not be assessable in a standardized format. Be careful with that as peer review clearly calls for all standards to be assessed.

2.x.RF.4.B Reading with automaticity is difficult to assessment on a large-scale assessment.

2.x.R.1.A Measuring pre-reading skills will be difficult as they are a strategy, not an actual skill to be measured.

2.X.W.4.B It is difficult to know how a student constructed known words, so I'm not sure how we would measure that they did so using letter sound knowledge.

2.X.R.1 We want to teach close-reading strategies, and we hope students use them, but we can't actually measure the use of close-reading strategies.

Standard 8--Independent reading and writing--can't be measured in a large-scale standardized context. Great goals, though.

Marianne Perie
Director, Center for Educational Testing and Evaluation

Review – OAS English Language Arts Standards (Public 3rd Draft)

I have examined the draft standards focusing on the question of whether the English Language Arts Benchmarks as written can be assessed. My review concentrated on Grade 3 and above, although the benchmarks at the lower grades are assessable, but would need more individual observation than used in typical Grade 3 and higher assessments.

The best way to describe the possibility of assessing the English Language Arts Standards is to examine the seven standards separately. Each requires different approaches. In the broadest sense all content is assessable. However, the question is really "How practical is it to assess standards in a statewide group administered assessment?".

Standard 1-Speaking and Listening

These standards require students to work with other students in demonstrating competencies. The speaking standards will also require a one on one observational assessment. Although this has been tried is some states it is very difficult to administer and collect data statewide when the assessment relies on such a large part on local administration and scoring.

Standard 2-Reading and Writing Process

These standards are focused on the lower grades and extend to Grade 5 in places. Other than some read aloud assessment which would have to be taped or individually administered the content is assessable with statewide assessment techniques which would include open-ended items or exercises.

Standard 3-Vocabulary

The standard for Reading can be assessed. The Writing can also be assessed and is probably more valid when assessed in an essay format.

Standard 4- Critical Reading and Writing

This standard describes the main content that is in present Reading and Writing assessments. For assessment to be valid there needs to be an open-ended component to the assessment. This is especially true in Writing where essay production is needed to properly assess the standards.

Standard 5-Language

These standards can be assessed in a largely multiple-choice assessment. In recent years these standards have been played down in favor of essay writing. The scoring of essays can only minimally serve to assess this standard.

Standard 6-Research

This standard seems to call for the student in Grades 6-12 to complete a project or paper using research methods. This is probably assessed best in the classroom not at a statewide level.

Standard 7-Multimodal Literacies

As in the Research standard this content is not easily assessed in statewide assessment. Assessment could be combined in a project with Standard 6.

Standard 8-Independent Reading and Writing

This standard is not assessable by statewide assessment and seems to be focused on student self reflection which would seem to be classroom based.

Standards 2, 3, 4, and 5 make up the statewide assessable content in these standards. Multiple-choice only assessments will not be ideal for some of the standard content. The approach of using open-ended questions, essays, and multiple-choice formats can provide assessment for these standards.

Review – OAS Mathematical Standards (Public 3rd Draft)

I have examined the draft standards focusing on the question of whether the Mathematical Benchmarks as written can be assessed. My review concentrated on Grade 3 and above, although the benchmarks at the lower grades are assessable, but would need more individual observation than used in typical Grade 3 and higher assessments.

In almost all cases I believe these can be assessed and have seen examples of items that assess this content as stated in other state assessments. My review assumes that the assessment allows for the use of some open-ended items and some graphical items that are interactive. If the assessment is limited to only multiple-choice items then assessing some of the benchmarks which require production of explanations will be more difficult and will depend on interpretation. These newer conceptual content benchmarks are almost always represented by open-ended questions.

I have inserted comments in a few places in the early grades where I thought the wording might confuse assessment.

The upper level Algebra and Geometry do call for the use of some tools (graphing calculators and computer type software like spreadsheets). These may be costly to provide in an assessment environment, because they also must be provided to the instructional setting for training. Before finalizing these standards it will be necessary to identify exactly what "appropriate tools" are for the assessment.

One caution about the "new content" contained in the standards. It is important that the content be both assessable and accessible. Some of this material is probably not being taught in all schools and it will be important to establish these standards and the terminology within before assessment. Please do not forget your first guiding principle-high standards and equity in the form of support for students.

John Keene- OK TAC Member
Assessment and Evaluation Services

Review of OAS ELA Standards for Assessibillty (Public Draft 3)

After examining the draft standards, I have concluded that while some of these standards may be assessed at the State level, some clearly cannot, and many will require a change in emphasis from selected-response items to constructed-response items. This will have multiple impacts upon the State system, including increasing costs of item and test development, reduced turn-around time due to the need to have rater-grading, and sufficient investment in computing facilities to handle the technology-enhance items that will be needed.

My quick glance suggests that many of the standards fall in line with Common Core standards, and research by PARCC, for example, has provided insight into how these standards may be partially assessed via large-scale testing using technology-enhanced items. Although many of these standards items could be assessed, the time and cost of doing so may not be of value to the State.

Here are some general observations. Any standard that requires the physical presence of another person to accomplish (mostly Pre-K through 2) will be difficult to assess on a large scale. Any standard that requires using additional idiosyncratic software to produce a response (e.g. multimodal literacy) is likely

to be difficult to assess. Any standard that requires any kind of idiosyncratic selection (e.g. a topic of the student's own choosing) will be difficult to assess. With that in mind, here are my thoughts regarding the standards.

Standard 1. Most of these in the early grades are not assessable. The first examples require the physical presence of others:

Reading:

1.7.R.1 Students will actively listen and speak clearly using appropriate discussion rules with awareness and control of verbal and nonverbal cues.

1.7.R.2 Students will ask and answer clarifying questions and acknowledge others' ideas presented orally, through text or other media.

1.7.R.3 Students will engage in collaborative discussions about appropriate topics and texts, expressing their own ideas clearly while building on the ideas of others in pairs, diverse groups, and whole class settings.

Writing:

1.7.W.1 Students will give formal and informal presentations in a group or individually, providing evidence to support a main idea.

1.7.W.2 Students will work effectively and respectfully with diverse teams, exercise flexibility and willingness to make necessary compromises to accomplish a goal, assume shared responsibility for collaborative work, and value individual contributions made by each team member.

Although these may be done at a local level, the psychometric rigor of these assessments is likely to be poor.

Standard 2. Many of these are accessible using compute appropriate software and constructed –response items. For example, computer programs currently exist for assessing phonological awareness, although technology investments (e.g. headphones) will be necessary. However, those requiring physical presence (i.e. moving finger to match voice, which would best be done using eye-tracking equipment) will be difficult to assess reliably and with a large-scale assessment . Plus, many of these standards cannot be assessed except for open-ended (CR) type items (c.f. **2.3.R.1.B** Students will ask and answer literal questions, using the text to support answers). Here are some other examples of difficult to assess standards using large-scale assessments:

Reading:

2.PK.RF.3.A With guidance and support, students will name the majority of the letters in their first name and many uppercase and lowercase letters. (Too idiosyncratic to the person)

2.K.RF.2.B Students will demonstrate correct book orientation and identify (by pointing) the title and the front and back covers of a book. (Requires physical observation)

2.K.RF.2.C Students will point to show that written words are made up of letters and are separated by

spaces.

2.K.RF.2.D Students will point to show that print moves from top to bottom, left to right, and front to back (does not have to be matched to voice).

2.3.RF.4.C Students will orally read grade-appropriate text at an appropriate rate, smoothly and accurately, with expression that connotes comprehension at the independent level. (Requires independent observation)

2.K.R.1.B With guidance and support, students will ask and answer questions about texts during shared reading.

Writing:

2.PK.W.1 With guidance and support, students will begin to use appropriate grip to hold a writing utensil when drawing or writing. (Physical observation – not compute-based)

2.PK.W.2 With guidance and support, students will write the majority of the letters in their first name and some uppercase and lowercase letters. (Idiosyncratic to first name)

2.4.W.6 Students will use resources to find correct spellings of words (examples: word wall, vocabulary notebook, print and electronic dictionaries and spell-check.). (Resource availability constraints)

Standard 3. **Vocabulary.** Most Vocabulary standards accessible using compute appropriate software and constructed –response items. However, those requiring physical assessment (i.e. read-alouds; cf **3.PK.R.1** With guidance and support, students will begin to develop an awareness of context clues through read-alouds and other text experiences) will be more difficult and will require more resource. Many of the writing standards require the production of writing and thus the expenditure of time and resources - not to mention turn-around time – and thus are of concern.

Standard 4. Critical Reading and Critical Writing. Early grades seem to suggest a lot of with guidance and support; can this be accomplished via a computer-program or does an actual individual need to be present? The latter will limit the assessability of the standards.

Questions that involve idiosyncratic responses will be difficult to assess in a large-scale assessment (cf. **4.2.R.4.A** Students will compare their own point of view with that of the narrator or characters in a text.).

Standards involving the authors intentions and purposes have been notoriously difficult to achieve consensus (see the Pineapple story and the Fable : NY Times: the author did not agree with the test developers interpretation) as to the authors meaning.

Background knowledge differences between ethnic groups may make certain standards difficult to assess (cf. 4.9.R.1 Students will evaluate the extent to which historical, cultural, and/or global contexts affect authors' stylistic and organizational choices).

Overall, much of the standard is best assessed via CR items (especially the writing standards).

Standard 5. **Language.** The vast majority of the language standards can be assessed via MC –items, using CR items only if desired. If you choose to understand the phrase "the student will compose" are requiring active composition, then CR items will be required. If the students can passively select from among various compositions that meet the requirements (e.g. adequate mechanics) then a fully MC-based

assessment could be constructed.

Standard 6. Research. Research simulation-type items have been constructed by PARCC, blending MC-type items with CR-type (prose-constructed) items. Although these capture the synthesizing aspect of the research process, they don't capture the active part of the finding the relevant information. Because many of the standards here are idiosyncratic to the student, they will be difficult to assess on a large-scale. Likewise, resource constraints and availability will make certain standards difficult to assess in a large-scale assessment. Examples include:

6.1.R.2 Students will consult various visual and text reference sources to gather information. (What resources will be available - just online?).

6.2.R.1 Students will use their own questions to find information on their topic. (Idiosyncratic)

6.2.R.2 Students will identify the location and purpose of various visual and text reference sources.

6.1.W.1 Students will generate questions about topics of interest. (Idiosyncratic).

6.1.W.2 With guidance and support, students will organize information found during group or individual research, using graphic organizers or other aids. (Group research?)

6.1.W.3 Students will make informal presentations of information gathered. (To whom and in what modality? Will we be taping them?).

6.7.R.1 Students will use their own viable research questions and thesis statements to find information about a specific topic. (Idiosyncratic).

6.7.R.2 Students will evaluate resources from both primary and secondary sources (print and/or digital). (Print? How will this happen in a short amount of time?).

6.7.R.3 Students will follow ethical and legal guidelines for collecting and recording information. (Will follow? Are students going to conduct a study, collect information, and record it in real-time? It is easy to know about the guidelines … it may be harder to follow them).

6.12.W.1 Students will integrate evidence by quoting, summarizing, paraphrasing, and citing sources to create projects and presentations for multiple purposes while avoiding plagiarism. (Cannot be assessed in a short time frame as in a test).

Standard 7. Multimodal Literacies. This content is very difficult to assess in a large-scale assessment. The writing standards seem to require the production of multimodal content, which would require lots of time, software standardization, and background knowledge (designing for global communities). I do not believe that many of these standards can be assessed at the State Level.

Standard 8. Independent Reading and Writing. Course Management Software (e.g. Desire-to-Learn) provides information about what, when, and where access to readings and when the production of writing occurs. So, using computer-assisted technology to track student activities – provided all meaningful activity occurs via computer - is certainly possible in the age of big data. However, it is not currently feasible.

Summary: Much of the standards are currently assessable and follow along the path of the Common Core assessments (See PARCC for example). Standards 1, 6,7, and 8 appear to be the least assessable

given current technologies and time and cost considerations. The clear movement toward active testing via constructed-response items as seen throughout these standards suggests that the common goal of quickly turning around results for instructional purposes will no longer be a priority.

OAS Math Standards Review

First, let me say that I appreciate the guiding principles presentation and found these useful while evaluating the Math standards. Many of these standards are Common Core consistent and item developments for these types of standards have already been released (PARCC, 2013). Although these standards seem much more assessable than those for the ELA, there are still some concerns that are raised.

First, the PARCC development seems to center around the use of technology-enhanced items and the use of many CR-type items for follow-up explanations. Thus, the use of equation editors, spreadsheets, graphical tools, and simulation software are taken as a given in the PARCC assessment and would be useful for the OAS Math standards as well. This raises several concerns about timing, cost of item development, and cost of technology. Standardization of software is important in these assessments, both at the time of assessment and during the instructional phase. So which spreadsheet, equation editor, graphing calculator and so on will be used? And will all Oklahoma students (following GP 1) have timely access to these tools?

Second, asking the why questions (which are good questions to ask) requires CR-type questions which invokes delays in scoring and reporting back to the teachers for instructional purposes. Since the State has requested a quick turn-around on these assessments, it is quite possible that little if any feedback will be returned at the pace Oklahoma teachers have come to expect.

Finally, some of these standards cannot be easily assessed in a large-scale assessment, if at all. For example, under Actions and Processes we find "Develop a Productive Mathematical Disposition" and "Develop the Ability to Communicate Mathematically". I have no idea how these will be assessed, especially the former. I suspect they won't be.

There are some other specific issues. Benchmark A1.D.2.3 (Calculate experimental probabilities by performing simulations or experiments involving a probability model and using relative frequencies of outcomes) involves performing simulations –which usually involve computers but sometimes random objects (dice) or random number tables. There are numerous programs and applets available via the WEB that can conduct these simulations for you – but you are not performing the simulation you are merely observing the results. I do not believe that students can be assessed on performing a simulation – it takes too long. Understanding the results of a simulation - perhaps.

Many of the benchmarks in Data and Probability suggest collecting data. Taking this literally, does this mean that students s have to actually collect data? Time and ethical considerations suggest that this part will not, should not, and cannot be assessed.

Summary: Most of the Math Standards are assessable using traditional MC-items, technology-enhanced items, and CR-type items. Time and cost considerations (at both the instructional and assessment contexts) are the most important factors affecting the assessibility of these standards, in addition to the increased turn-around time.

Robert Terry. Ph.D.

Professor, University of Oklahoma