

8/26/2017

To: Shaun Bates and Lisa Sireno

From: Andy Porter, Chair, Missouri Technical Advisory Committee

Subject: Minutes of Missouri TAC Meeting on August 17 and 18, 2017

The Missouri Technical Advisory Committee met at the Renaissance St. Louis Airport Hotel from 10:00 AM to 5:00 PM, August 17, 2017, and from 7:30 AM to 12:30 PM, August 18, 2017. Members of the TAC in attendance were Bertha Doar, director of assessment, St. Louis Missouri Public Schools; Karla Egan, independent consultant; Ron Mertz, St. Louis Public Schools, retired; Barbara Plake, University of Nebraska, retired; Andy Porter, chair of TAC, University of Pennsylvania; Ed Roeber, independent consultant; and Phoebe Winter, independent consultant; In attendance from DESE: Lisa Sireno, Shaun Bates, John Kitchens, Commissioner Margie Vandeven, Deputy Commissioner Stacey Preis, Assistant Commissioners Blaine Henningsen, and Chris Neale. In attendance from DRC: Lindy Wienand, Joanna Tomkowicz, Sara Brazzle, Rick Mercado. In attendance from Questar: Adam Johnson, Mike Woods, Sandra Durden, Katie McClarty and Scott Bishop. In attendance from MetaMetrics: Ellie Sanford-Moore.

Contractors did not join the meeting until lunch on the first day.

### **Missouri update**

Lisa Sireno provided the TAC with an update on assessment activity in Missouri. The Missouri Show Me Standards provide process and content standards for the state. The Missouri Learning Standards address Grade-Level and course content. New fine arts standards will soon be drafted, but no assessments of them are planned at the present time. The personal finance standards will soon be revised and a new assessment is planned for 2019/20. Missouri adopted new learning standards for mathematics, English language arts (ELA), social studies and science in 2016. The Missouri peer review submission to the U.S. Department of Education was largely approved for Algebra I, English II and grade-level Science as well as MAP-A in mathematics and ELA.

MAP-A continues to be provided by Dynamic Learning Maps (DLM) for grades 3 through 8 and 11 in ELA and mathematics and grades 5, 8 and 11 in science. The state continues regular assessments in ELA and mathematics for grades 3 through 8 and in science for grades 5 and 8. Data Recognition Corporation (DRC) is developing interim assessments for Missouri. End-of-Course (EOC) testing continues. Sireno noted that this past year, approximately 20% of the eighth grade students took the Algebra I EOC.

Missouri continues to use the WIDA consortium for assessing English language competence. For the past three years the state has done census assessment using the ACT for 11<sup>th</sup> graders. Recently the state cut \$4 million from the assessment budget (out of the approximately \$18 million budget provided by the state, supplemented by \$7 million from the federal government).

The current plan is to address this shortfall by discontinuing the state-funded ACT census assessment at 11<sup>th</sup> grade.

There will be new assessments for Grade-Level testing in ELA and mathematics and the EOC assessment in 2017/18. There will be new assessments for fifth and eighth grade Science and the EOC assessment in 2018/19 and in social studies for 2019/20 but only in high school.

DESE has published several resource documents for state educators concerning the new standards.

Sireno cited as challenges to assessment in Missouri the frequency and size of changes to state standards and the impact on assessments. The number and frequency of changes has created a great deal of confusion among K-12 educators in the state. There have also been challenges to the EOC assessments, a topic of discussion for later in the meeting. At the end of the conversation about challenges, the TAC noted that many of the materials for the current meeting were received too late to be read in advance and that this practice needs to be corrected.

### **Grade-Level assessments technical report early tables for 2016 /17**

Materials were received in advance of the meeting and Joanna Tomkowicz from DRC walked the TAC through the results for ELA, mathematics and science. In addition to reporting percent of students at various proficiency levels and mean scale scores, there were also results on interrater reliability and differential item functioning (DIF). When investigating subgroup differences, there continue to be large differences between average performance for black students and white students and general education students versus students with disabilities. Tomkowicz noted that no items were dropped from use based on the DIF analyses. DESE reminded the TAC that all items were reviewed in advance of their initial use by appropriate state content committees. **The TAC wondered whether it would be useful to consider bringing in some “advocacy groups” to look at DIF flagged items.** The reliabilities of the assessments for subgroups were investigated and reported as good to excellent despite the possibility of restriction in range for some of the subgroups in terms of their performance on the assessments.

**The TAC thanked Tomkowicz and DRC for the quality of their work and was pleased to see that the Grade-Level assessments continue to function well in Missouri.**

### **Grade-Level assessment data forensics**

Reports were circulated to members of the TAC in advance of the meeting on analyses of answer changes, test time and item time. In addition, there were analyses reported of the possibility of students copying answers from one another. Overall there was very little evidence of results flagged as suspicious and possibly needing investigation except for test time. Tomkowicz noted that for the forensics analyses of time, the amount of time for an individual student is not necessarily the real time because students didn't always remember to log out.

**The TAC did note that the reports refer to flagging in terms of numbers of standard deviations when in fact the flags were identified in terms of number of standard errors.**

**This should be corrected.** In the future, DRC will investigate the possibility of unusually large gains in student achievement from one year to the next. These results will also be analyzed by crossing them with wrong to right answer change analyses.

**The TAC complimented DESE and DRC for their attention to data forensics in detecting the possibility of cheating. Further, this commitment and the resulting analyses are surely serving as a deterrent to any cheating in the state. Sireno reported that DESE is developing follow-up plans in response to data forensics flags of classrooms and schools. The TAC requested that they have an opportunity to review these follow-up plans when available. The TAC recommends that the forensics analyses also look to see if there are any school districts that receive a surprising number of flagged classrooms/schools and that data analysis be added to the current analyses. The TAC noted that flagged/classrooms might sometimes be an indication of administration problems and those possibilities should be investigated as well.**

### **Embedded field tests in spring 2017**

In spring 2017, 393 ELA items and 612 math items were field tested. Each item was taken by approximately 3,000 to 4,000 students. For ELA, 73 items were flagged largely for low or negative correlations with total score. For mathematics, 203 items were flagged primarily because they were too difficult. All flagged items were reviewed by DESE employees and dropped or modified as appropriate.

### **Multi-select items**

Multi-select items are in a multiple-choice format, but instead of only one right answer, several of the answers provided below the item stem can be correct and appropriately chosen. While there is not a great deal of evidence in support, some believe that these multi-select items can test content in the standards that is difficult to test with other item formats. A sheet of rules for writing these types of items was distributed at the meeting, and discussed in detail; revisions were suggested by the TAC. Generally, the TAC favored a single set of rules to apply to all three subjects: ELA, math and science. Further, the TAC favored simplifying the rules where possible.

### **Providing math formula sheets with the assessments**

Thus far in Missouri assessments, sheets providing a variety of mathematical formulas that are accessible to the student while taking the exam have not been provided. Nevertheless, there are some who would prefer that formula sheets be provided. The TAC asked how the state content standards address this issue and was assured that the standards were silent on the matter. **The TAC recommends that the state continue to not provide formula sheets for mathematics assessments. If there should come a time when an item is written which is meant for the formula to be provided, that formula can be provided in the item itself. If that happens, care should be taken the formula provided does not change what is tested by any of the other items on the test form.**

### **2017/18 ELA and mathematics vertical scaling plan**

The TAC was provided in advance of the meeting an email from Joanna Tomkowicz to Lisa Sireno and Shaun Bates describing DRC's plans for vertical scaling for the spring 2018 MAP assessments in ELA and mathematics developed to be aligned to the new Missouri learning standards. DRC proposes to use a set of common items between spring 2017 and spring 2018 operational tests in every grade ELA and mathematics to serve as anchor items in developing the new vertical scales. In addition, sets of items from the grade above and the grade below will be administered to samples of students taking on Grade-Level operational tests to facilitate between- grade assessment linking. A hybrid approach to building vertical scales as recommended by the TAC in a previous meeting will be used. In this approach, concurrent calibration results are treated as an initial scale the operational assessments. The initial scales are then equated to the existing ELA and mathematics scales using the common anchor items between the two administrations. Concurrent calibration of test data for all grades and content areas will then be conducted to build the vertical scales.

The metric of the new vertical scales will be established to have means and standard deviations substantially different from the previous scales to signal that these are new assessments for new standards and comparisons of results to earlier years should not be made.

**The TAC complimented DRC on their plans for building the new vertical scales. The TAC did recommend that vertical linking items be carefully selected so that they reflect the best available items for spanning alignment of content between the two adjacent grades. The TAC would like to see the results for building the vertical scale using a) items from above and below grades, b) items from only above grades and c) items from only below grades before deciding which approach should be used.**

### **Interim assessments**

DESE has asked DRC to develop an interim assessment for grades 3 through 8 in ELA and mathematics and for grades 5 and 8 in science to be ready for the 2017/18 year. These interim assessments consist of one form per subject and grade and are also being called "pretests." The TAC asked what the goals were for these pretests and was told they would be used as practice tests that mimic the operational tests in format and length. The practice tests could be taken any time and as many times as desired between the period November and June. The plan is to calibrate the practice test to the operational test using first-time test-taker data only. Results would be reported on the same scale as the operational test and end of year proficiencies would be reported.

**The TAC liked the plans, but recommends that the tests be called "practice tests," not "interim assessments" nor "pretests."**

### **Vertical articulation between middle and high school assessments**

DESE has asked its two contractors (DRC and Questar) to work together to better articulate high school assessments and Grade-Level assessments when new standards are set for each. The two contractors jointly talked to the TAC about their plans, saying that a formal proposal for vertical

articulation will be presented to DESE and the TAC at the next meeting. The goals of the vertical articulation are to have a K-12 assessment system where the proficiency impact data is coordinated between the assessment systems for K-8 and high school. Further, where there are jumps in the impact data, the articulation should provide an explanation for why those jumps make sense. Essentially, the idea is that if a student is proficient at one Grade-Level he or she is ready to be proficient at the next higher Grade-Level.

With those goals in mind, the two contractors plan to set high school standards first. Maybe they will use ACT benchmarks and impact data to inform the high school standard-setting. Then, Grade-Level proficiency standards would be set within the context of the previously set EOC standards. At one point in the discussion, a rationale for higher percents proficient for EOCs was offered, noting that the focus of the course and the assessment for a single course might cover a narrower domain than the domain for an entire Grade-Level for a subject. These narrower domains might be more easily taught and mastered. **The TAC noted that the contractors use different IRT models and that this might influence the selection of a response probability for standard-setting. The TAC further noted that there can be difficulties for standard-setting when a test is centered on the standards but not well centered on actual student achievement levels. The TAC concluded that while it is all right to have different response probability values for grades 3-8 and high school, the same response probabilities should be used across subjects and grades for grades 3-8 and across courses for End-of-Course tests in high school. The TAC was informed by DESE that they are working on a policy piece addressing these issues; and the TAC recommends that that policy piece be included in the proposal from the two contractors addressing vertical articulation of the two assessment systems.**

### **EOC 2016 /17 results**

DESE Commissioner Margie Vandeven joined the TAC for a discussion of whether the 2016/17 results for Algebra I and English II are valid for use in the state accountability program. The discussion was a continuation from the TAC conference call held in July. The concern is that student performance in these two subjects dropped substantially from the previous year. In its July conference call, the TAC concluded that this was a form effect, at least in part, rather than a true drop in performance. The TAC noted that when the same form was used two years prior as was used in 2017, performance was comparable (although somewhat higher). While it is common to use different forms over years for EOC tests, the forms were not built to ensure that they are psychometrically on the same scale. Importantly, 2017 is the last year for using the non-equated forms; in the future, EOC assessment forms will be carefully built so that the forms are not only reliable and valid but also comparable. At this point in time, results of the 2016/17 assessments have been reported to students, but their use for accountability has not yet happened. The interpretation of results of the other EOC assessments is not being called into question.

The TAC was informed that a performance index is built based on student proficiency levels with 1 point for below basic, 3 points for basic, 4 points for proficient and 5 points for advanced. Using this index, three measures are used for accountability: progress, value-added and status. The TAC noted that neither value-added nor progress requires that forms be comparable between 2016 and 2017 since value-added is metric free and progress is measured such that it reflects

only change between two years ago and the current year (i.e., 2015 and 2017); for those two years the same form was used. The problem then is status.

Discussion identified six possible actions:

1. Drop the 2016 results and use only 2015 and 2017 for which the same form was used.
2. Hold schools and districts harmless, meaning that if a school or district does better in 2017, on results as reported they get credit for the improvement, otherwise they get the same accountability levels as for the prior year.
3. Cut scores could be adjusted by, say, a single raw score point to make the results between 2016 and 2017 appear more comparable.
4. Use equipercentile procedures, putting the form used in 2017 on the same scale as the form used in 2016
5. Accept the results as reported to students and use the results for accountability accordingly.
6. Exclude the results for Algebra I and English II in accountability and don't report those results.

**Of the above six possibilities, five members of the TAC favored the sixth option, not using the results for accountability and not reporting them. Two members of the TAC favored the first option; drop the 2016 results and use only 2015 and 2017 for which the same form was used. By the end of the meeting, however, all seven members of the TAC favored the sixth option. The TAC recommends that in the future, every effort be made to analyze the data before the results are reported to students so any problems can be detected and corrected.**

### **Small-scale pilot update**

Questar is developing new writing prompts and performance tasks. In the previous meeting, the TAC recommended that these be pilot tested and in a later TAC meeting, the TAC reviewed plans for a pilot test. At the time of that review, TAC strongly recommended that the pilot be online as is the operational test. At the current meeting, Questar reported that the pilot will take place on September 3; the pilot will be online and approximately 200 students reflecting a range of achievement levels will be used to pilot each prompt and performance task. There will be a different sample of students for each prompt and task. **The TAC complimented Questar on these plans.**

### **Stand-alone field test for science**

Questar plans to build four forms each for Biology and Physical Science: two operational forms, one pre-test form, and one breach form for each assessment. Apparently, 95% of the students in the course will take the field test. **There was some discussion as to whether or not students would be comparably motivated to do well on the field test as they will be motivated on the**

**new operational test. There was some discussion as to whether or not items in the field test could include items from the item pool to estimate the size of motivation effects. Ultimately, the TAC was supportive of Questar’s plans for the field test.**

### **Lexile® and Quantile® measures**

Ellie Sanford from MetaMetrics led the TAC in a discussion of how Missouri might report student achievement in reading in Lexiles® and in mathematics in Quantiles®. DESE is interested in exploring these possibilities and asked that the TAC provide their thoughts on the pros and cons of doing this. The motivation for the initiative is to provide students and parents with new ways of thinking about student achievement levels and appropriate future instruction. These reporting metrics are based on the concept of prerequisites. Knowing what prerequisites students have mastered provides suggestions as to where their next instruction should be.

Sanford mentioned that one possibility was to embed MetaMetrics items in Missouri assessments so that the Missouri assessments could be reported on the MetaMetrics scales. Discussion revealed that currently many districts in Missouri are using other tests than those given by the state that report results in Lexiles® and Quantiles®. **The TAC wondered whether it would be possible to use these data to link to state assessment results and put the state assessment results on the MetaMetrics scales rather than having the additional student burden of adding items to the state assessments to accomplish that end.**

**The TAC was uncertain as to DESE’s questions concerning MetaMetrics and its products. DESE will prepare questions for the TAC to address on these issues at a future meeting.**

### **In closing**

The next meeting of the TAC is scheduled for December 7 and 8, 2017, and the meeting after that for March 8 and 9, 2018. Yet a third meeting was scheduled for August 2 and 3, 2018.