Assessment Design Considerations for the MAST Pilot

July 6, 2022

Note to the Task Force: We tried to summarize and synthesize your good work from June 20-21 regarding the various considerations for the design of key assessment system components. Throughout the document that follows, you'll find highlighted questions for you. Please think about these questions and your responses in preparation for our meeting on Wednesday.

The MAST Task Force spent considerable time at its June 20-21, 2022 meeting translating its robust theory of action into specific design considerations for the innovative assessment. The Task Force focused on the following major assessment components:

- Item Specifications
- Assessment Specifications
- Administration Requirements
- Score Reporting

For each of these major assessment components, we present the Task Force's recommendations regarding assessment design as well as the rationale tied to the theory of action.

Item Specifications

The focus of the discussions related to item specifications was on the types of test items presented to students as well as the general framework for presenting the items. The Task Force recommended including a variety of item types on the various through-year assessment events including:

- Selected-response
- Evidence-based constructed response questions
- Short constructed-response (scored automatically) No longer than a short paragraph.
- Technology-enhanced items

The Task Force's rationale for the different item types was based on the tradeoff between what the educators could learn about student knowledge and skills. Task Force members noted that the constellation of item types could vary by content area and the age/grade of the students tested. The Task Force felt that multiple-choice items were more familiar and generally more accessible to most students, but the various types of open-response items and tasks better allow for the

measurement of more complex and authentic thinking than is the case with multiple-choice items. The Task Force was particularly interest in exploring the use of technology-enhanced items (TEIs) because such items offer the promise of advancing measurement capacity in a cost a cost-efficient manner.

Questions for the Task Force

1. What's the balance among item types for the through-year assessment events?
2. Should the items all be on grade-level? Is your answer the same no matter what time of year the assessments are administered?

Assessment Specifications

The discussions about assessment specifications addressed several key areas of assessment design including:

- Delivery platform
- Delivery approach
- Nature of the assessment events
- Grain size of the through-year components

We discuss each of these aspects of assessment design below.

Delivery Platform

The Task Force strongly recommended computer-based testing with the full range of accommodations for students with identified disabilities and English learners similar to what is currently offered with the Smarter Balanced. The Task Force also recommended using a standardized method of identifying students who need the accommodations and/or using a universally designed approach to ensure that all students are able to access the assessment to the fullest extent possible. The Task Force agreed that a limited number of paper-based tests should be available for students who are unable to access the assessment via computer.

Delivery Approach

The Task Force discussed the opportunities and limitations associated with computer-adaptive compared with fixed-form approaches for presenting test items to students. Computer-adaptive tests (CAT) are those where the difficulty of each item presented to a student varies depending on the student's responses to the previous items. The test adapts to maximize the information gathered for each student as efficiently. Of course, this is the ideal case. In practice, the efficiency is constrained by item sampling requirements from subdomains and test length.

A fixed-form test, on the other hand, is one where all students in a given grade and subject area see essentially the same set of test items. This also varies somewhat because of field test and matrix-sampled items used for equating purposes.

In general, the Task Force recommended using a CAT approach to the extent practical, but some Task Force members thought it could be beneficial to use fixed form tests if item analysis reports could be provided to educators. However, item analysis reports requires releasing a considerable number of test items which has serious cost implications.

Questions for the Task Force

1. The Task Force needs to clarify its recommendation for CAT vs. fixed form for the TY components.

Modular or Mini-Summative

Modular designs are those that are tied to specific and discrete learning targets expected to be taught at various times throughout the year. Mini-summative designs are those where each test event (e.g., fall, winter, and spring) is meant to sample the knowledge and skills expected to be learned by the end of the school year. This is how the current state summative assessment is designed, except a through-year mini-summative design follows this blueprint (the map of knowledge and skills expected to be tested) each time the test is given.

The Task Force strongly recommended employing a modular design for the through-year components. The specific design of these modular assessments, in terms of the standards and/or

other groupings of the content and skills each was expected to assess, should differ across content areas and likely across grade levels as well. The Task Force strongly recommended having Montana content experts and educators closely involved in the design of the overall system and each through-year assessment.

Questions for the Task Force

1. Is this an accurate representation of the discussion?

Administration Requirements

The Task Force discussed two aspects of through-year test administration: the frequency with which through-year assessments would be administered and the sequencing of such administrations. The Task Force wanted some ability to measure fall-to-spring growth but did not settle on a specific number of through-year assessments to include in the system as long as there were at least three.

Like the recommendations regarding the type of design, the Task Force wanted the maximum flexibility possible in terms of the sequencing of the through-year assessments.

Questions for the Task Force

1. The desire for measuring fall-to-spring growth contradicts the desire to have modular assessments. If growth is desired, it is best to have the two tests as similar as possible. That said, there are some models to measure "growth" across modular assessments, but it would require a common sequence for at least the 1st and last test of the year. The Task Force will need to weigh in on these tradeoffs.
2. Some members of the Task Force indicated a desire for having teachers select which modular assessments to administer depending on their local curriculum sequence. We (the Center) recommend NOT allowing these decisions at the teacher level. Rather, any flexibility should be at the district level only. What does the Task Force think of this recommendation?

Score Reporting

Score reporting is likely the most important aspect of assessment design because it is the only way that assessment owners communicate about the assessment to multiple sets of stakeholders. As such, the Task Force discussed access to the various reports and then offered initial thoughts on reporting at the student, parent/community, educator, and leader levels. However, for all reports and for the system as a whole, the Task Force emphasized the need to support substantial assessment literacy and related professional learning opportunities. These learning opportunities should include much more than simply explaining how to interpret the reports, but also how to use the assessment results to further the achievement and growth of participating students.

The Task Force strongly recommended that all users have access to timely, effective data regardless of location or economic status. This included ensuring that all users have sufficient internet capability. This means that schools must have adequate bandwidth and enough devices so that all educators and students can access the reports.

Student Reports

The Task Force recommended that the student reporting portal include a dashboard that presents within and across year longitudinal growth. The dashboard should also include a comprehensive set of scores and goals related to these scores.

Parent Reports

In addition to what is presented on the student reports, the parent report should emphasize how parents are able to connect to school learning and life connection and engagement in student learning. The parent reports should also emphasize school-wide and perhaps district-level performance. These reports should allow for the easy monitoring of student growth and achievement.

Classroom Reports

The classroom reports include all of the information as the student and parent reports but must include more detailed information, particularly fine-grained information related to specific

learning targets (or standards). In particular, the educator reports should present the cumulative picture of student achievement and growth developing throughout the year.

The Task Force indicated that it would be helpful for the classroom reports to offer instructional suggestions such as grouping students, identifying particular resources that might be useful for bridging learning gaps, and recommendations for enrichment planning. The Task Force emphasized the importance of facilitating educators' flexible use of the data to first correctly interpret the results and what they mean in terms of each student's knowledge and skills. Second, teachers' access to the relevant information should enable them to plan appropriate instructional actions.

School and District Reports

The Task Force recommended that school and district leaders are supported in accessing a rich set of information via a performance dashboard that includes achievement on each of the through-year assessments, summative determinations, within- and across-year student longitudinal growth broken down by content areas, grade levels, student groups, classrooms, and intersections of the multiple categories.

The Task Force noted that in addition to presenting comprehensive information, great care should be taken to design the various score reports to maximize the utility for each of the targeted user groups. To the extent possible, the score reports should include suggested actions for designated user groups, but this might not be possible given flexibility in administration and other sources of uncertainty.

The Task Force strongly recommended having the data backbone meet key interoperability standards to facilitate data use for local comprehensive school improvement planning (CSIP), as well as state and federal reporting requirements.

Questions for the Task Force

1. Did we miss anything important related to our discussion of score reporting?

2. Someone mentioned that they would like the reports to present information to support within-state (e.g., at the LEA and/or school levels) as well as national comparisons. The national comparisons are certainly not possible with the flexibility of the system desired, and within-state comparisons could be challenging as well. We should discuss this more.