



# Assessment Design and Implementation Considerations for the Montana Alternate Student Testing (MAST) Pilot Program

Final Report From the Montana Alternate Student Testing Task Force

Elsie Arntzen, Superintendent of Public Instruction  
Montana Office of Public Instruction



AUGUST 2022



# TABLE OF CONTENTS

## TABLE OF CONTENTS

THE TASK FORCE .....	3
TASK FORCE PROCESS AND MEETINGS .....	4
THROUGH-YEAR ASSESSMENT SYSTEMS .....	4
• Additional Goals of Through-year Assessment.....	5
THEORY OF ACTION .....	6
• Introduction to Theories of Action .....	6
• The MAST Theory of Action.....	6
ASSESSMENT DESIGN CONSIDERATIONS FOR THE MAST PILOT .....	7
• Item Specifications.....	7
• Assessment Specifications .....	7
• Administration Requirements.....	8
• Score Reporting .....	8
MAST PILOT IMPLEMENTATION RECOMMENDATIONS.....	9
• Timeline for Pilot and Full Implementation .....	10
• Technology Infrastructure.....	10
• Administration Guidance.....	10
• Assessment Literacy and Professional Learning.....	10
• Role of Grant Schools with Required Interim (and Other) Assessments.....	11
• Impact on Small-Schools .....	12
• Communication with Stakeholders .....	12
• Designated Project Leadership .....	12
• The Continuing Role of the Task Force.....	12
CONCLUDING THOUGHTS .....	12
APPENDIX A: MONTANA'S THEORY OF ACTION .....	13

REPORT PREPARED BY:  
SCOTT MARION, PH.D.  
JUAN D'BROT, PH.D.  
W. CHRIS BRANDT, PH.D.

*The National Center for  
the Improvement of  
Educational Assessment*

The Montana Office of Public Instruction (OPI) is exploring the design and piloting of a “through-year” assessment system. OPI contracted with the National Center for the Improvement of Educational Assessment, Inc. (Center for Assessment) to assist in developing a framework for its potential innovative assessment design. Toward this end, the Center for Assessment facilitated a task force of Montana stakeholders to develop a theory of action that can help frame the design of the new assessment system and identify priorities for OPI to consider for pilot implementation.

This report outlines the framework for state assessments and informs the process of transitioning to a next-generation assessment system. Specifically, this report describes the work of the task force and presents the theory of action for guiding the overall design and evaluation of the system as well as the implementation of the pilot. This report addresses both major issues.

## THE TASK FORCE

OPI leadership invited diverse stakeholders to participate in the task force. Although many potential task force members could not participate for various reasons, OPI nonetheless recruited a knowledgeable and representative set of stakeholders, as seen below.

FIRST NAME	LAST NAME	AFFILIATION	STAKEHOLDER CATEGORY
Crista	Anderson	Dixon	Principal
Norah	Barney	Anaconda	Special Education Specialist
Scott	Beagles	Libby	Curriculum Director
Opal	Besaw	Kalispell	Student
Jonna	Brandt	Missoula	Curriculum Director
Jilyn	Chandler	Helena	Elementary School Principal
Ashley	Davis	Gallatin Gateway	High School Teacher
Rick	Duncan	Deer Lodge	Superintendent
McCall	Flynn	Helena	State Board of Education
Sandra	Francis	Hardin	System Test Coordinator
Jane Lee	Hamman	Clancy	State Board of Education
Drea	O'Donnell	Sheridan	District Superintendent
Angie	Nelson	Savage	Middle School Principal
Carol	Paul	Great Falls	English Learner Specialist
Pam	Pettigrew	Glendive	Special Education Specialist
Karen	Pollari	Sidney	Elementary School Teacher
Shaun	Scott	Helena	Higher Education/School Board
Beth	Walsh	East Helena	Middle School Teacher



In addition to the Center for Assessment staff, the task force was supported by staff and leadership from OPI.

## TASK FORCE PROCESS AND MEETINGS

Between May 4th and July 28th, the Center for Assessment facilitated four virtual 3-4 hour webinars and one 2-day in-person meeting with the task force. These meetings involved identifying perceived problems with, as well as things task members liked about, the current assessment system, outlining common assessment design principles, drafting a theory of action, and describing considerations for implementing the pilot. Additionally, we explored the potential innovative design put forward by the New Meridian Corporation and OPI.<sup>1</sup>

The Center for Assessment had worked from a preliminary theory of action developed by the task force, along with their announced likes and dislikes with the current system, to produce a detailed draft theory of action which the task force reviewed and refined at the June 20-21 in-person meeting. Additionally, task force members identified critical considerations for the assessment system design and pilot implementation.

The July 6th webinar focused on reviewing and revising these design and implementation considerations and, further, discussing other issues related to the draft theory of action. The Center for Assessment used this information to produce a draft report for review and feedback at the July 28th webinar, the result of which is this final report.

## THROUGH-YEAR ASSESSMENT SYSTEMS

Several states, including Montana, are in various stages of exploration, design, and development of a through-year assessment system, where:

- multiple, distinct assessments are administered throughout the school year, and
- these assessments are intended to support both (a) a summative determination for each student and (b) at least one additional goal.

In other words, a through-year assessment system involves a distributed design that provides a summative determination of student proficiency on state content standards as required by current federal law and, further, supports a more specific goal (e.g., identifying students needing extra support). A summative determination can take the form of a total score or an achievement level (e.g., proficient) at least once per year, generally at the end of the year.

Current interest in through-year assessment systems appears to be based on strong assumptions about the multiple purposes that such assessments may serve. These assumptions, in turn, lead to claims about the types of inferences and uses the assessment system is intended to support. Notably, a theory of action reveals the logic underlying these claims and inferences, and it also specifies the resources and processes necessary to support the goals of the through-year system.

### *Current Examples of Through-Year Assessment Systems*

Many people regard the administration of three or four independent assessments as the typical through-year design, consistent with how districts administer commercial interim assessments. There are many design variations, however, including:

- weekly, standards-aligned assessments based on a diagnostic classification modeling system;
- delayed-stage adaptive systems, in which through-year assessments administered during the year serve as “priors” for the end-of-year assessment;
- unit-based assessments tied to a common curriculum; and
- modular assessments tied to specific groupings of content standards.

OPI is proposing a modular design, where each module is tied to important groupings of knowledge and skills represented by the content standards, which districts (or perhaps schools) choose to administer in the order that best fits the scope and sequence of the district’s curriculum.

<sup>1</sup> This design was the subject of the proposal OPI submitted to the U.S. Department of Education’s Competitive Grants for State Assessments.



## ADDITIONAL GOALS OF THROUGH-YEAR ASSESSMENT

Crafting a successful through-year assessment system requires prioritizing goals, acknowledging that not all goals can be met within a single system. The first step in specifying the additional goal(s) is to identify the problem users are trying to address that led to their interest in a through-year design in the first place. This was a key focus of the task force.

### *What's the Problem?*

In our work, the Center for Assessment has encountered five general concerns with the typical end-of-year test, making a through-year assessment system particularly attractive. First, many users, including district and school leaders, teachers, and even parents, have expressed concerns that their current assessments do not produce instructionally useful information. Most assessment experts acknowledge that end-of-year tests are not designed to provide fine-grained instructional information, but that does not prevent users from wanting more.

Second, most school districts already administer interim assessments throughout the year, such as NWEA MAP, Renaissance STAR, and Curriculum Associates i-Ready. The last interim assessment is usually given in the spring. This often leads to concerns about double testing, where an interim and state summative test occur in close proximity. Of course, it is the district leader's choice to double-test, but that does not stop them from pushing for the elimination of the state test because they argue they are getting the information they already need.

Third, some state leaders would like to design and implement an end-of-year test that includes rich performance tasks, both to better measure the depth of learning and to signal the types of activities teachers should structure in their classrooms. Unfortunately, administering such tasks takes considerable time. State leaders are reluctant to face stakeholder pushback due to excessive testing time and are concerned about the associated validity threats due to student fatigue. Thus, they may turn to a through-year design to spread out end-of-year testing.

Fourth, education leaders and teachers may be concerned that assessing students just once each year fails to capture the student's true level of knowledge and skills because of test anxiety, personal factors, and other idiosyncrasies associated with single-occasion testing. Therefore, these stakeholders claim that offering students multiple assessment opportunities throughout the school year provides for more accurate measurement of students' knowledge and skills.

Finally, the cancelation of state testing in the spring of 2020 because of the COVID-19 pandemic, revealed the single point of failure of state testing. Specifically, stakeholders had placed so much emphasis on end-of-year test results that, when the test disappeared, they felt they were left with no data. Having a set of assessments aligned with state content standards and administered throughout the year would have avoided this predicament.

### *Additional Goals*

Again, a through-year assessment system must address at least one additional goal beyond federal accountability. All assessments included in the through-year system are intended to support accountability purposes and at least one other use. These additional goals derive from the concerns described above. Some common goals associated with current through-year systems are:

- State leaders hope to provide instructionally useful information to educators and students throughout the year to enhance the value proposition of state tests beyond accountability. The design could include a flexible schedule for administering through-year assessments to better align with the scope and sequence of the district's curriculum.
- State leaders hope to reduce the burden of a single-test administration. The end-of-year summative assessment could be spread out so that less time is required per testing occasion (although not necessarily less total testing time). This was the motivation behind the original through-year design of the Partnership for Assessment of Readiness for College and Careers, one of the U.S. assessment consortia funded through Race to the Top in 2010.<sup>2</sup>
- State leaders hope to develop a more coherent assessment system than afforded by the current combination of an end-of-year state summative assessment with various off-the-shelf interim assessments. Through-year designs arguably strengthen the alignment of interim and end-of-year state tests than currently is the case.

The additional goals should inform the claims that the assessment system is designed to support. These claims, in turn, inform the validity argument necessary to evaluate the inferences derived from the assessment system's scores. For example, test designers may claim that a through-year assessment system will provide results that can be used to address the instructional needs of students. In this case, scores from through-year assessments must produce information at the appropriate grain size and timing to support valid instructional decisions. This is the rationale for the interest in a through-year system in Montana, and it is the one we explore in the following section.

<sup>2</sup> This design was never operationalized because of concerns regarding the impact on curricular sequencing in participating states and the time requirements for multiple administrations.

# THEORY OF ACTION

## INTRODUCTION TO THEORIES OF ACTION

A theory of action is a critical tool for designing and evaluating policy. Theories of action are schematic representations of the logical argument connecting the long- and short-term outcomes of a policy initiative to the processes and mechanisms intended to affect those outcomes. A theory of action clarifies alternatives and potentially competing claims about how a policy initiative should work. Policymakers and designers must explicitly outline how the proposed design choices are intended to accomplish the goals of the through-year system. In other words, why is the pilot designed in the way it is to fulfill the intended goals?

Further, policymakers also must describe the *how*. For example, *how* will involving local educators in the design of the assessment system influence teacher knowledge and student learning? Having to articulate both the aims and mechanisms of the program may expose proposed policies for evaluating schools that are untenable and, further, will shed light on some fruitful means of meeting the major policy goals. For example, placing too much accountability weight on the through-year components will likely interfere with any instructional utility.

When outlining the theory of action, the task force had to ensure that connections among various aspects of the assessment system are not simply belief statements. Rather, these connections should be supported by research, or, if research is not available, by a strong theoretical argument backed by best practices. The links between the elements of the theory of action should serve as testable hypotheses to be verified with pilot-implementation evidence. This accumulation of evidence would support the ongoing validation of the assessment system.

## THE MAST THEORY OF ACTION

The MAST theory of action is shown in Appendix A. As noted above, the task force began developing a theory of action by outlining goals for a new assessment system. Next, the task force proposed the resources, activities (inputs), and actions (processes) required to realize short-, mid-, and long-term outcomes. This is a complex initiative, and it consequently requires a detailed theory of action to guide the design, implementation, and evaluation of an innovative assessment system.

The MAST theory of action depicts two major aims (i.e., the primary and additional goal) of the through-year assessment system: providing instructionally useful information and supporting comparable annual determinations. Also, notice the important role of communication and stakeholder engagement that the task force deemed important for the success of the program. In the case of the proposed pilot, the instructional goal is intended to be supported by the flexible administration of modular assessments tied to local curriculum—and at a grain size small enough to support useful instructional actions. Further, score reports must be designed to facilitate clear and accurate interpretations of students' knowledge and skills relative to the identified learning targets.

The next four panels of the theory of action highlight key details of the theory of action. The first detailed panel, shown in Figure 2, focuses on the conditions, processes, and intermediate indicators necessary to increase the quality and utility of feedback provided to students and teachers from the through-year assessments. This panel emphasizes the professional learning and structures necessary to support classroom-based instructional uses.

Figure 3 addresses technical issues regarding the design of the test items and test forms, as well as the technology infrastructure necessary to support the proposed innovation. Outcomes from these technical enhancements include a higher-quality assessment and a more engaging assessment experience.

Figure 4 focuses on the connections among standards, instruction, assessment, and professional learning. The intended outcomes of this aspect of the theory of action include improved educator practices, particularly regarding how teachers provide feedback to students and opportunities for deeper engagement in subject matter.

Communication is the focus of the last panel (Figure 5). The task force was emphatic that absent a clear and compelling communication strategy, the pilot likely would fall short of its goals. Further, the task force believed OPI must lead all communication efforts so that a consistent statewide message is delivered. OPI also should provide support to district leaders for communicating effectively with local stakeholders. The intended outcomes of this component of the theory of action include a clear understanding of the pilot by all stakeholders and high levels of buy-in.

# ASSESSMENT DESIGN CONSIDERATIONS FOR THE MAST PILOT

The task force spent considerable time at its June 20-21 meeting translating its robust theory of action into specific design considerations for the innovative assessment. The task force focused on four major assessment components:

- item specifications
- assessment specifications
- administration requirements
- score reporting

For each component, we present the task force’s recommendations regarding assessment design. These recommendations were based on the task force members’ interpretations of how best to operationalize the theory of action.

## ITEM SPECIFICATIONS

Here, the discussion focused on the types of test items that students will experience as well as the general framework for presenting these items. The task force recommended that a variety of grade-level item types be included in the various through-year assessment events:

- selected-response
- short constructed-response (scored automatically)
- technology-enhanced items

The corresponding rationale was based on the tradeoffs regarding what 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. They believed that selected-response items are more familiar and generally more accessible to most students, but that open-response items are better for measuring more complex and authentic thinking. The task force was particularly interested in exploring the use of technology-enhanced items because of the items’ cost-efficiency in measuring complex skills and abilities.

## ASSESSMENT SPECIFICATIONS

The discussion of assessment specifications addressed several key areas of assessment design:

- delivery platform
- delivery approach
- nature of the assessment events (modular or mini-summative)

We discuss each 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 assessment system. The task force also recommended using a standardized method of identifying students who need accommodations and using a universal design approach to ensure that all students can 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 versus fixed-form approaches for presenting test items to students. With computer-adaptive tests (CAT), the difficulty of an item presented to a student depends on the student’s responses to the previous items. In principle, such tests maximize the information gathered for each student as efficiently as possible. Of course, this is the ideal case. In practice, CAT efficiency, as with any traditional test, is constrained by item subdomain sampling requirements—the minimum number of items needed within a subdomain—and test length.

In a fixed-form test, on the other hand, all students in a given grade and subject area are given 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 members thought it could be beneficial to use fixed form tests if item analysis reports could be provided to educators. However, item analysis reports require releasing a considerable number of test items, which has significant cost implications. Further, the task force noted that not all modules need to be configured the same way.

### ***Modular or Mini-Summative***

Modular designs are tied to specific and discrete learning targets expected to be taught at various times throughout the year. With mini-summative designs, 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.

The task force strongly recommended employing a modular design for the through-year assessment components. The specific design of these modular assessments, in terms of the standards or other groupings of the content and skills each is 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.

## **ADMINISTRATION REQUIREMENTS**

The task force discussed two aspects of through-year test administration: the frequency with which through-year assessments would be given, and the sequencing of administrations. The task force was interested in measuring fall-to-spring growth if it was possible given other design constraints. Additionally, the task force thought it was important to have at least three unique assessments.

The task force wanted the maximum flexibility possible in terms of sequencing the through-year assessments. To aid in school and district planning, however, the task force recommended that district leadership decide the timing and sequence of the administration of these modules.

While some members wanted the ability to measure fall-to-spring growth, all members recognized the challenges of doing so with a modular design and flexible administration schedules. Ultimately, the task force prioritized the flexible administration of modules over fall-to-spring growth.

## **SCORE REPORTING**

Score reporting arguably is the most important aspect of assessment design insofar as it is the only way that assessment leaders communicate about the assessment to multiple sets of stakeholders. The task force discussed options for accessing score reports and offered initial thoughts on reporting at the student, parent/community, educator, and leader levels. Additionally, the task force emphasized the need to support assessment literacy and related professional-learning opportunities. Such opportunities should include not just how to interpret the reports, but also how to use assessment results to improve student learning.

The task force strongly recommended that all users have access to timely, effective data regardless of location or economic status. To accomplish this goal, members understood that schools must have sufficient bandwidth and electronic devices so that all educators and students can access the reports.

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 will depend upon administration sequences and local curriculum choices.

### ***Student Reports***

The task force recommended that the student reporting portal include a dashboard presenting within- and across-year longitudinal growth. The dashboard also should include a comprehensive set of scores and corresponding actions that might be taken.



### ***Parent Reports***

In addition to including what is presented on the student reports, the parent report should emphasize how parents are able to connect to school learning and make life connections that engage them in their student's learning. The task force recognized this might be an overly ambitious request, but they thought it was worth asking. The parent reports also should emphasize school-wide, and perhaps district-level, performance. These reports should permit easy monitoring of student achievement and growth.

### ***Classroom Reports***

The classroom reports contain the same 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, classroom reports should present the cumulative picture of student achievement and growth throughout the year.

The task force believed it would be helpful for classroom reports to offer instructional suggestions, such as grouping students, identifying resources to bridge learning gaps, and/or support enrichment opportunities. The task force emphasized the importance of facilitating educators' flexible use of the reports to (a) correctly interpret the results and what they mean in terms of each student's knowledge and skills, and (b) given these data, plan appropriate instructional actions.

### ***School and District Reports***

The task force recommended that school and district leaders have easy access to a rich set of information via a performance dashboard that includes achievement on each of the through-year assessments, summative determinations, and both within- and across-year student growth—all broken down by content area, grade level, student group, classroom, and intersections of the multiple categories (low income by special education status).

The task force strongly recommended having the data backbone meet key interoperability standards to facilitate data use for local comprehensive school improvement planning, as well as for satisfying state and federal reporting requirements.

## **MAST PILOT IMPLEMENTATION RECOMMENDATIONS**

After considerable work on the draft theory of action and the assessment design considerations for the MAST pilot, the task force, at their June 20-21, 2022 meeting, began formulating recommendations for the successful implementation of the pilot. These recommendations were grouped into the following major categories, with several topics comprised of assessment literacy and professional learning:

- timeline for pilot and full implementation
- technology infrastructure
- administration guidance
- assessment literacy and professional learning
  - score interpretation and use
  - curriculum and instruction guidance and professional learning
  - leadership support
  - interim assessment guidance
- role of grant schools with required interim (and other) assessments
- small school impact
- communication with stakeholders
- project leadership
- the continuing role of the task force

For each major category, we present the task force's recommendations regarding pilot implementation, again, based on the task force's interpretation of the theory of action.

## **TIMELINE FOR PILOT AND FULL IMPLEMENTATION**

The task force recommended a phased pilot, beginning with limited implementation in 2022-2023 in grades 5 and 7 in both ELA and mathematics. Rather than recommending a full pilot, which would provide enough students to generate reliable item statistics, the task force instead recommended more of an item tryout from which OPI and its assessment partners can learn whether the modular assessments are working as intended and teachers are able to use the results productively. For example, this limited pilot could involve trying out various configurations of the modules (e.g., regarding length, standards covered, grain size). Task force members also suggested using the 2022-2023 school year to conduct increased outreach to the various stakeholder audiences regarding their perceptions of the current assessment system and their desires for an innovative approach. Further, the task force recommended exploring ways to best incentivize schools to participate in the pilot, such as relief from other state requirements. This limited pilot could then be expanded to grades 4 and 6 the following year, with the potential for full implementation in 2024-2025.

The task force also recommended collecting data on implementation successes and challenges across the pilot experience. Data collection activities should inform recommendations regarding professional development and other considerations affecting pilot implementation. Further, the task force recommended that OPI clearly communicate the expectations, and provide clear guidance, for all participants in the process. Task force members believed this level of transparency is crucial for building support for the pilot and, ultimately, the innovative assessment program itself.

## **TECHNOLOGY INFRASTRUCTURE**

Task force members were concerned that all Montana schools did not currently have the technology infrastructure (e.g., bandwidth, devices) necessary to participate in the pilot. Therefore, OPI and Montana legislature need to provide the resources necessary for all schools to successfully engage in the innovative assessment system without causing a disruption to teaching and learning when the same technology is needed for both assessment and instruction. Therefore, the task force recommended that OPI and its contractors conduct an audit of technology capabilities for all schools (unless one has already been conducted), and, in turn, evaluate the results of the audit against the demands of the innovative system. To be sure, technology shortfalls will compromise the pilot's success more than almost any other aspect of this endeavor.

## **ADMINISTRATION GUIDANCE**

The task force recommended that OPI and its partners make the administration rules and conditions (e.g., test security, test windows, settings) exceptionally clear before launching the pilot, and then clearly communicate these rules and expectations to all stakeholders. Pilot schools must be provided with adequate professional learning opportunities to ensure the successful administration of the assessment modules. Professional development should focus on the big picture, such as the pilot's goals, and intended uses of through-year data.

Additionally, the task force recommended that OPI, in consultation with participating pilot schools clarify how the pilot, specifically the modular assessments, are intended to interact with the district's existing assessment system. As we discuss below, this could include recommendations that districts pause, or limit, their current use of interim assessments to avoid over-testing and mixed messages. Task force members asked that pilot schools be granted relief from the current state summative assessments during the pilot. Unfortunately, such relief from the federal testing requirements could be granted only if OPI applied for, and received, flexibility as part of the ESSA Innovative Assessment Demonstration Authority (IADA). Thus, the task force suggested that OPI apply for the IADA as soon as the application re-opens.

## **ASSESSMENT LITERACY AND PROFESSIONAL LEARNING**

The task force stressed the importance of providing extensive professional-learning opportunities associated with various dimensions of the pilot and at various levels of the system. Members suggested establishing and funding regional education centers that could more regularly provide professional development, technical support, responsive answers to questions, and other forms of outreach to local schools throughout the process. Further, rather than having professional development come solely from the State, the task force suggested that the State partner with district and school leaders to ensure that local education stakeholders are responsible for professional learning. In the following subsections, we present task force recommendations regarding specific types of professional-learning needs.

### ***Score Interpretation and Use***

The theory of action offers considerable detail about the importance of accurate interpretation of the various test scores produced by the innovative assessment system. The instructional utility of the pilot system requires that educators and leaders take appropriate actions, informed by test scores, to improve student learning. If these stakeholders misinterpret test scores, their subsequent actions are compromised.

The task force noted the importance of designing high-quality score reports to make interpretation as easy as possible. The task force identified several types of data they would like to see on score reports, such as whether students have mastered certain knowledge and skills, and in what areas students need instructional support. Most importantly, the task force recommended engaging in a deliberative process regarding score-report design to ensure that the various stakeholders' needs are well-understood.

Task force members suggested automating standard classroom reports that flag students, student groups, or entire classes that require targeted support. The task force also suggested developing a tool, such as an app, that allows educators to tailor reports on an as-needed basis (although some of this functionality is already built into many applications).

### ***Curriculum and Instruction Guidance and Professional Learning***

The task force acknowledged the strong local control ethic in Montana regarding curriculum, personnel decisions, instructional approaches, and the like. Nevertheless, the task force recognized that, for this pilot to be successful, OPI may need to offer more professional-learning opportunities and guidance than is normally the case. Specifically, the task force suggested that OPI evaluate and report on the alignment, in terms of both sequencing and the targeted knowledge and skills, between the through-year assessments and the major curriculum packages used in Montana. This evaluation should provide insights regarding instructional best practices related to Montana content standards. Task force members also suggested engaging the Montana Council of Deans of Education so that the entire strategic plan begins to come together across constitutional and statutory entities.

### ***Leadership support***

The task force recognized that the success of the through-year pilot requires sufficient buy-in from school and district leaders. OPI and the project personnel therefore will need to provide professional-learning opportunities for education leaders so they feel confident in their capacity to serve as the assessment leaders in their schools and districts.

These professional-learning opportunities must be more than simply pilot-administration training. Rather, professional development must focus on helping leaders interpret assessment results and, in turn, use that information to guide and support educators in making appropriate instructional decisions. As district and school leaders increase their efficacy in applying innovative assessment resources, they can more successfully motivate teachers to engage in the pilot and in their own professional learning.

### ***Interim assessment guidance***

The task force considered whether OPI should provide district guidance regarding their current interim assessment programs. For example, OPI could issue such guidance as, "OPI suggests that schools/districts participating in the new assessment system should curtail the use of its current interim assessment system to avoid over-testing and potential mixed messages." The task force noted that while such guidance might be reasonable, there are many schools committed to interim assessments for a variety of reasons, including those related to the use of the Elementary and Secondary School Emergency Relief Fund (ESSER). Therefore, the task force recommended against offering formal guidance to schools and districts regarding the use of interim assessments during the pilot period. OPI may want to reevaluate this position once the state successfully administers the new system statewide.

## **ROLE OF GRANT SCHOOLS WITH REQUIRED INTERIM (AND OTHER) ASSESSMENTS**

Several task force members mentioned that "grant schools" are required to use scores from a commercial interim assessment program to monitor, and report on, their progress. Grant schools receive funds as part of the Montana Literacy Grant program, the Montana Transformational Schools initiative, and the Montana Multi-Tiered System of Support grant program. The task force recommended allowing grant schools to substitute the scores from the pilot initiative for commercial interim assessment scores. On the other hand, task force members recognized that schools participating in one of these grant programs would be unlikely to participate in the pilot because they would be reluctant to give up existing trend information based on their current interim assessments.

## IMPACT ON SMALL-SCHOOLS

The task force noted that small schools—of which there are many in Montana—would need extra support, resources, and incentives to join the pilot. For example, several of the concerns discussed above, such as technology infrastructure and professional capacity, are exacerbated in small schools. The task force also stressed that additional methods would be particularly helpful for evaluating the efficacy of the pilot in small-school contexts, such as conducting interviews and focus groups.

## COMMUNICATION WITH STAKEHOLDERS

The task force emphasized that OPI must clearly communicate the rationale for this pilot and the intended outcomes, which is to improve student learning at scale. “We need to share with our districts the ‘what and why’ of this process,” task force members noted.

The task force strongly recommended using multiple channels to provide information to all stakeholders. The task force thought it was critical for all state and local board members, superintendents, principals, and teachers, through their respective professional organizations, to receive early and regular communication about the pilot. Further, the task force recommended collecting student, teacher, and parent feedback during and after the pilot and, in turn, transparently sharing the results with all stakeholders. The task force emphasized that OPI must communicate before, during, and after innovative assessments are administered.

## DESIGNATED PROJECT LEADERSHIP

The task force was concerned that the pilot likely would fall short if OPI lacked sufficient capacity to manage the project. Therefore, the task force recommended that OPI hire, or otherwise designate, a project leader for the pilot. This leader should be the primary point of contact for district leaders and other pilot stakeholders.

## THE CONTINUING ROLE OF THE TASK FORCE

While it sounds self-serving, the task force strongly recommended that they, or a group like them, be established to provide advice to OPI and its contractors regarding the implementation process.

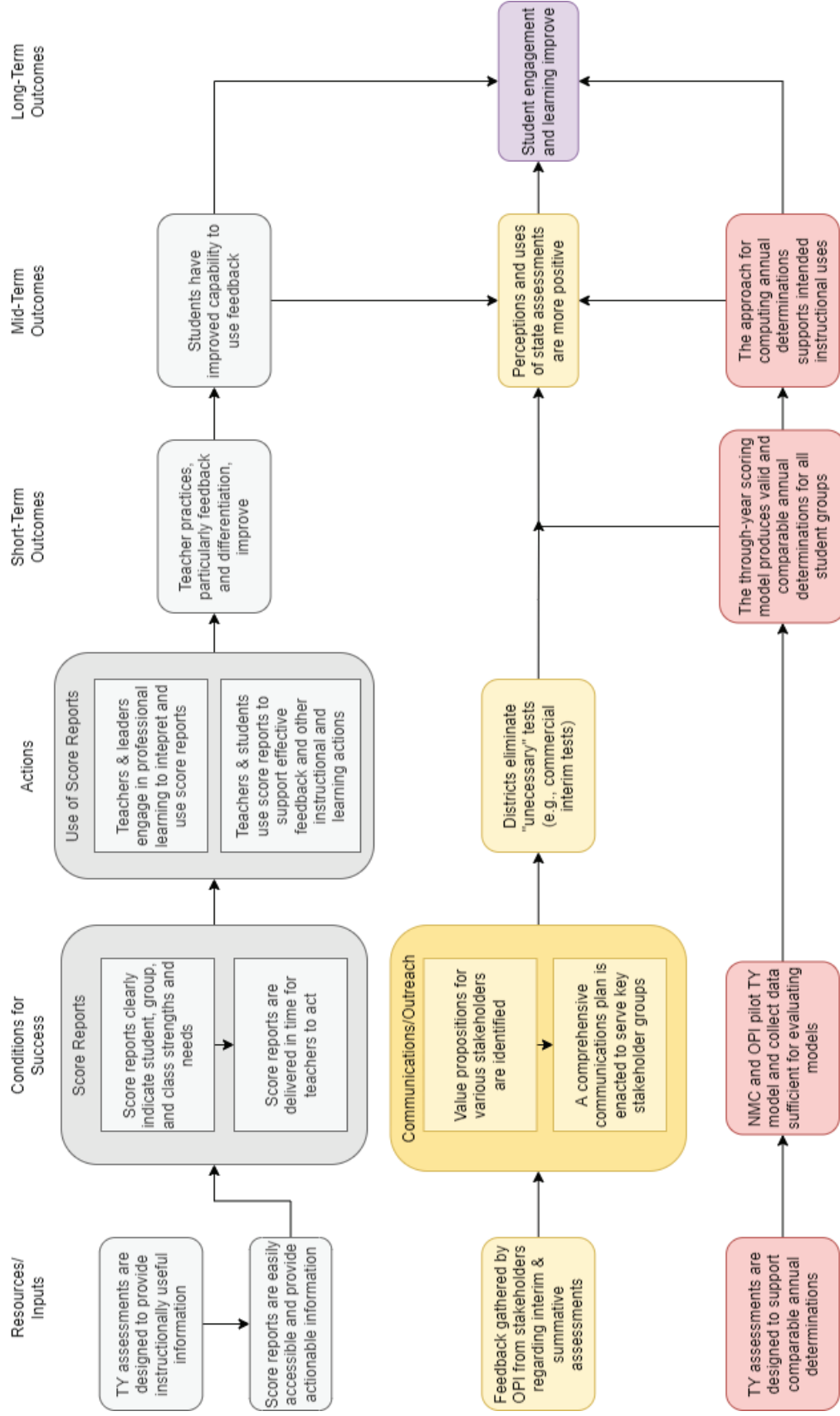
Further, the task force recommended that more students be added as formal members, along with test coordinators, school counselors, and perhaps others. The task force recommended establishing regular meeting cycles to monitor progress and provide feedback on the innovative assessment system and associated pilot program.

## CONCLUDING THOUGHTS

The MAST task force engaged in serious deliberations regarding the design and implementation of a through-year assessment system in Montana. They worked through the development of a detailed theory of action that describes the inputs, processes, and near- and long-term indicators necessary for the successful implementation of a through-year assessment system pilot. Additionally, the task force used this theory of action to describe some of the key assessment design considerations as well as offering recommendations for the successful implementation of the pilot. Task force members hope that OPI leadership finds the information contained in this report useful for guiding the design and development of Montana’s innovative assessment system.

# APPENDIX A: MONTANA'S THEORY OF ACTION

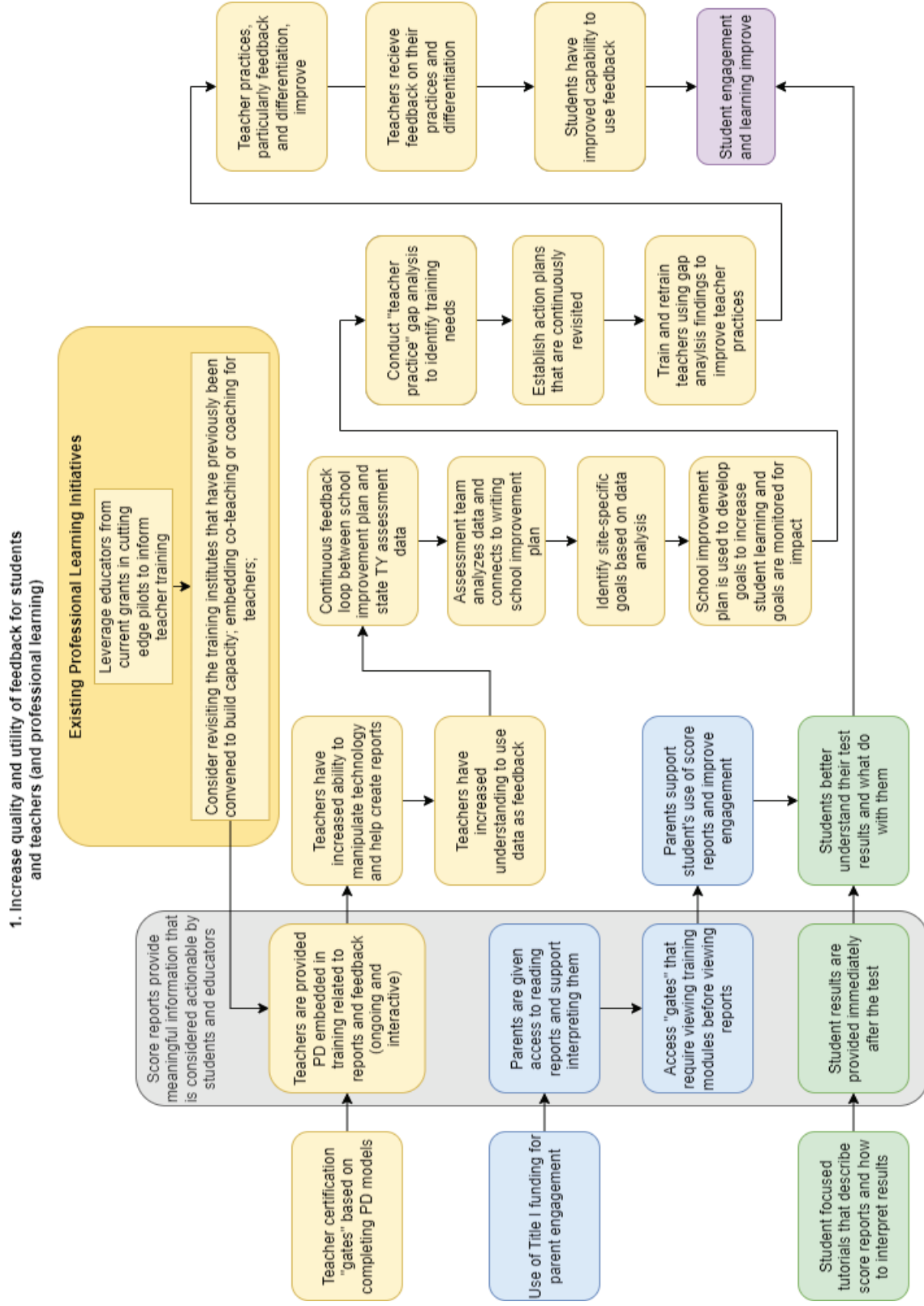
FIGURE 1. OVERALL THEORY OF ACTION



\*NMC = New Meridian Corporation; OPI = Office of Public Instruction; TY = Through-year



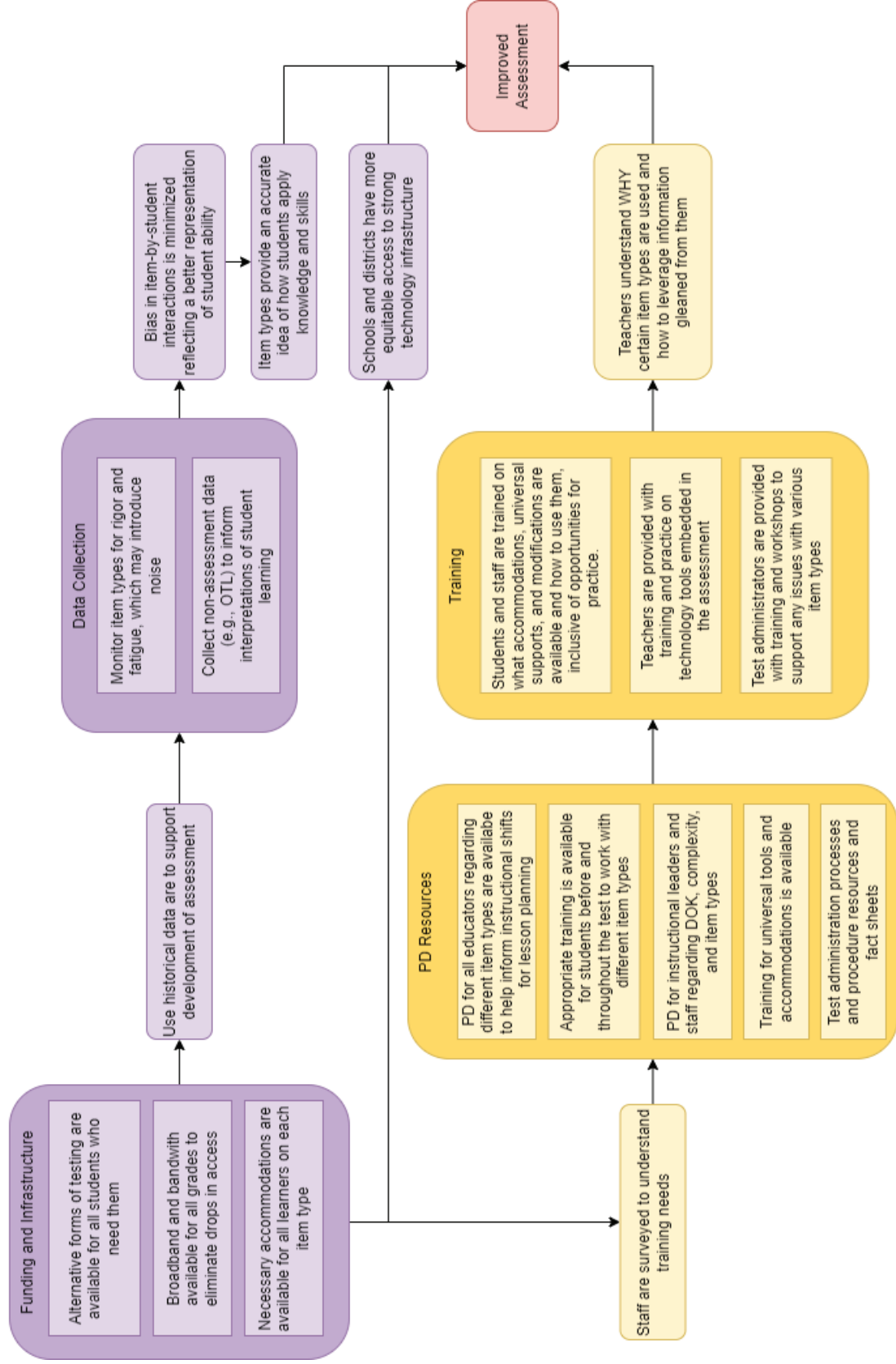
FIGURE 2. INCREASING THE QUALITY AND UTILITY OF STUDENT AND TEACHER FEEDBACK



\*PD = Professional development; CSIP = Continuous School Improvement Plan

**FIGURE 3. USING MULTIPLE ITEM TYPES AND TASKS TO BETTER REPRESENT STUDENT KNOWLEDGE AND SKILLS**

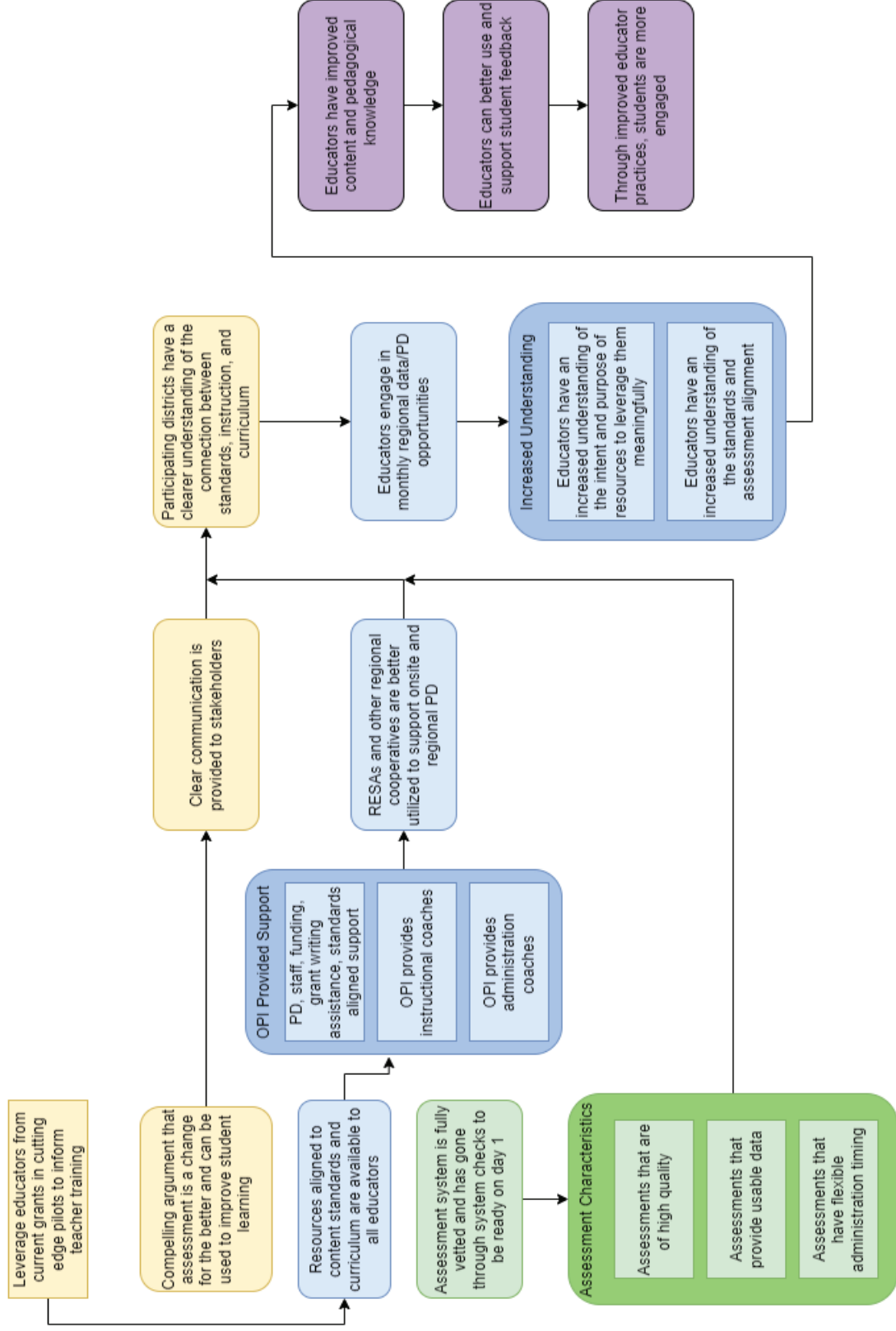
2. Ensure the use of multiple item types and tasks to better-represent student knowledge, skills, and performance levels



\*PD = Professional development; DOK = Depth of Knowledge; OTL = Opportunity to learn

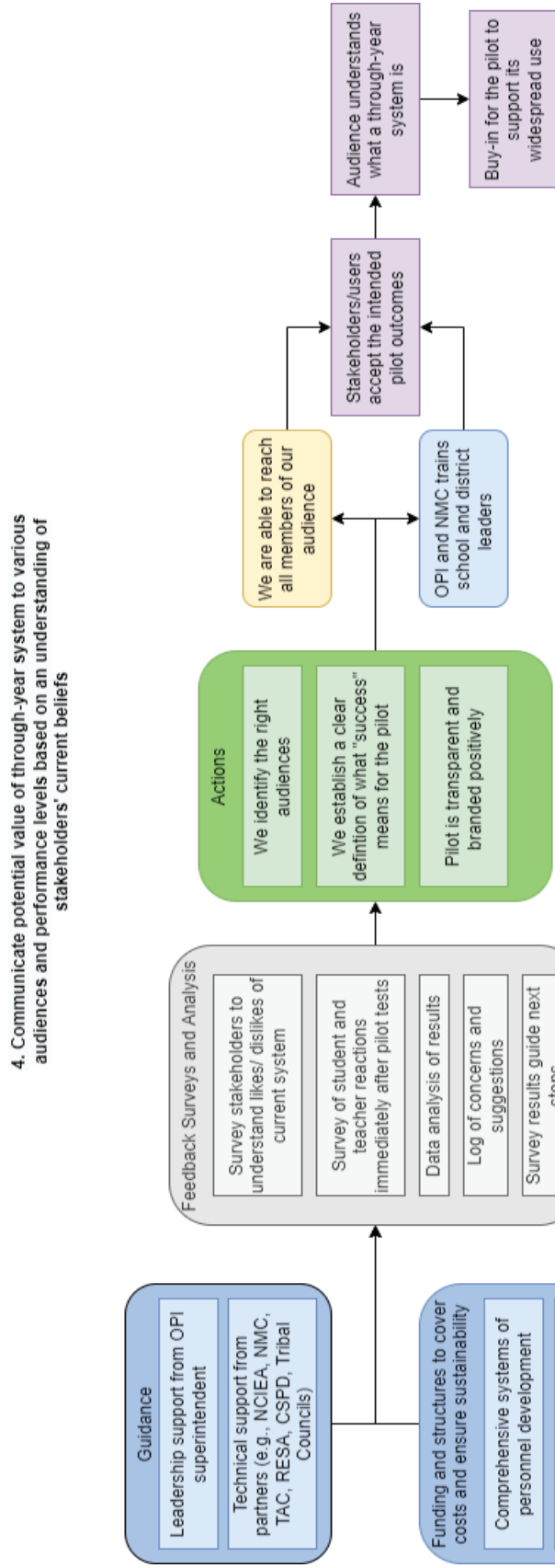
**FIGURE 4. ENSURING THE SYSTEM REPRESENTS A CLEAR CONNECTION AMONG STANDARDS, INSTRUCTION, AND ASSESSMENT**

3. Ensure new system represents clear connection among standards, instruction, assessment (and PL)




\*PD = Professional development; OPI = Office of Public Instruction; RESA = Regional Education Service Agencies

**FIGURE 5. COMMUNICATING THE VALUE OF THE THROUGH-YEAR SYSTEM TO A VARIETY OF AUDIENCES BASED ON THEIR CURRENT BELIEFS**



\*NCIEA = National Center for the Improvement of Educational Assessment; NMC = New Meridian Corporation; TAC = Technical Advisory Committee; RESA = Regional Education Service Agency; CSPD = Comprehensive Systems of Personnel Development; PM = Program/Project management; OPI = Office of Public Instruction



Assessment Design and Implementation  
Considerations for the Montana Alternate  
Student Testing (MAST) Pilot Program

AUGUST 2022

