

Feedback on Montana's Every Student Succeeds Act Accountability Plan

prepared by

Juan D'Brot

Jennifer Dunn

Introduction

This document has been prepared to provide some high-level feedback on the Montana state accountability plan. We approach this task as ‘critical friends’ to help provide objective advice and suggestions to support Montana’s commitment to developing strong, technically defensible solutions to support their policy priorities.

Our intent is to help the state establish a process for establishing overall goals for the educational system and an accountability system that helps them achieve these goals. In reviewing this document, we focused primarily on accountability and reporting [long term goals, challenging academic standards and assessments, and accountability], as these areas draw on our experiences and areas of expertise. Issues related to funding, school improvement, and teacher effectiveness were beyond the scope of our review.

Our feedback is presented informally as a series of bulleted notes we hope to talk through with state leadership. In general, as you approach refining the plan, we encourage you to think about the story you want to tell and articulate how the accountability system has been designed to help you tell the story. The following notes are placeholders and do not stand-alone without the context and explanation we hope to offer in a discussion.

Below, we present two sets of comments. First, we present comments specific to Montana’s draft Every Student Succeeds Act (ESSA) State Plan. Second, we present some general comments in developing Montana’s Plan and System. We would be happy to answer any questions Montana may have or provide additional thoughts or comments based on our recommendations.

Comments Specific to Montana’s Plan for Identifying the Vision, Goals, and Design Principles

In refining the overall plan, we suggest considering the following processes and specific recommendations:

- Establish a clear statement of the state’s policy priorities and values
 - For example: Montana prioritizes graduation from high school, promoting college/career readiness, equity, and developing non-academic skills in students.
 - In order to align with stated priorities and goals, long term goals should include a plan for establishing goals for college/career readiness and developing non-academic skills in addition to the stated goals for performance (General and ELL), growth, and graduation.
 - Clearly articulate how Montana wishes to conceptualize school quality. For example, by deepening the measure of student learning or by expanding the concept of school quality (e.g., through the use of improvement strategies or measures of student success).

- Clearly articulate long-term goals and how they are related (if at all) to how schools are classified annually.
 - The plan proposes performance and improvement goals centered on statewide averages. It is unclear whether these averages are based on proficiency rates, percent of students in each performance level, scales scores, or some indexed version of performance data. It will be important to
 - Articulate how the different assessments will be used in calculating the average (SBAC, ACT, MSAA, MontCAS). Will the goals be assessment dependent or will they be combined in some fashion?
 - Montana may want to revisit the long-term goal timelines. A three year timeline ending in 2020 might be a more reasonable short-term or interim goal. We recommend establishing a reasonable aspirational goal with an associated timeline that doesn't conflict with measuring progress toward attaining the goal.
 - Despite the impending overturning of the regulations, there is still a need to define statewide goals (informed by historical trends) as part of the ESSA plan . School-specific goals are not required to be based on statewide goals. However, it does make sense to establish school-level goals that are coherent with the statewide goals, but connected in a way that does not over identify the schools labelled as needing comprehensive support.
- Articulate the design principles being used to support these goals:
 - What are the desired features and characteristics of the system (e.g. can be differentiated, ability to disaggregate, can be standardized, makes use of reliable measures)?
 - How do you ensure that the design features are represented given the constraints of the data (e.g., standardization, data quality, data source)?
- Describe how the indicators conform with the design principles to support the goals of the system (e.g. subgroup performance, college and career readiness)
 - The following are mentioned as key factors at the start of the plan but then not in any of the detailed discussions. If they are to be included in the accountability plan, then more details need to be provided:
 - Measures of college and career readiness
 - Measures of non-academic skills
 - With regard to establishing long-term goals for the indicators themselves, this is not a requirement of ESSA outside of achievement, graduation rate, and English language proficiency. If the state wishes to include additional or alternative goals for indicators, this can become complex depending on how school-specific goals are related to statewide goals. How are these goals factored into school ratings, if at all? We are not advocating including indicator-based goals as part of schoolwide ratings. If, however, there is of interest to do so, be careful that the

indicator goals aren't in conflict with statewide goals and that the indicator goals do not create a trigger to over-identifying schools.

- For the EL goals, are the increases from baseline (i.e., 45%) based on historical increases in data or research? The advisory group may have referenced information that provides a rationale for this level of expected improvement. If so, it would be good to reference this to proactively describe to stakeholders the reasoning behind the increases. Additionally, we recommend comparing the desired increases to the WIDA Consortium data that has been compiled on ACCESS 2.0 to evaluate whether or not the increases are attainable.

Comments Specific to Montana's Plan for System Indicators and Aggregation

In refining Montana's accountability system for annual differentiation, we suggest considering the following processes and specific recommendations:

- Provide a clear explanation of how the indicators are being measured (e.g. percent proficient, growth percentiles, credit accumulation, percent of chronic absenteeism).
 - We recommend adding additional details of the specific calculations that will be conducted when you outline the measures. For example, how will ELL progress be defined? How much change will be deemed acceptable?
 - More generally, specify the metrics associated with each of the indicators. It is recommended that the quantification approaches balance the ability to communicate what each indicator means to the public with the ability to differentiate schools. For example, proficiency rates may not result in sufficient differentiation, especially with the more challenging Smarter Balanced assessments. If this is the case, it may be more appropriate to use the percent of students in each performance level with some form of weighting. However, this would require clear communication to stakeholders (especially educators and parents).
- For academic achievement it is unclear whether the state plans on using scale scores, proficiency rates, performance level counts/indices, or some other metric. In addition to the technical quality of the measure, the selection of the achievement measure should also consider the information you are trying to convey to stakeholders.
 - If the state opts to use scale scores in the system, it should be approached cautiously. Scale scores, let alone changes in scale scores can be very difficult to understand and interpret because there is no comparison for educators to reference. Additionally, it is easy to lose sight of proficiency expectations when using scale scores. Thus, it may be more important to include some link to goals when using scale scores or conscientiously report percent proficient alongside scale scores.
- With regard to growth, the state should consider the information gained from a model and the technical qualities of the model. General recommendations are made in the next

section of the document, but given that Montana’s plan alludes to the use of growth rates; we present two specific recommendations:

- Assuming the stakeholders wanted to expand the conceptualization of academic performance, the progress indicator should be compared to the academic indicator to ensure there is a sufficient degree of unique information that is provided. That is, the growth indicator should not be too correlated to the academic achievement indicator to help maximize differentiation in the classification system. Rate changes, especially if articulated as trajectories toward proficiency, may be highly related to academic achievement. However, if the stakeholders wish to make that a priority, then the weighting should be developed so that achievement and growth aren’t overloading the system.
- Additionally, examine the stability of growth rates (e.g., best fit lines) to determine the volatility in growth rates of schools over time. Estimates of rates may not be very stable until there are multiple (i.e., 4 or 5) years of data to establish a growth rate. Other growth approaches may be more stable and provide similar inferences of student growth.
- For the School Quality/Student Success indicator, School climate has emerged as an indicator of interest in a few states for ESSA planning. We present a few ideas to consider:
 - The consideration of how school improvement plans are being implemented or are viable is critical to the successful use of climate data. While climate data may be a measure of interest, it may be premature to use it as a high-stakes measure. This is based on the need to have an explicit theory of change or action at the school level to effectively use these data to drive change, which is reliant on the supports, resources, and conditions on the ground level (e.g., quality of the school’s leadership team or school culture to collect, interpret, and act upon data).
 - The acts of collecting, interpreting, and acting upon data (e.g., climate data) in a systematic fashion can be made explicit through school improvement plans that are focused so they are not arduous, but encourage schools to think through the links between the data, behavioral changes, and eventual outcomes.
 - The consideration of school improvement plan viability is naturally tied to this, but quantifying this with rubrics should be tested to ensure the process can be done systematically, can be replicated, and is defensible without the need for an unreasonable amount of resources (e.g., staff and time).
- The state should provide additional details of how the measures are being combined in a way that is supported by the measurement, consistent with the design principles and aligned with the overall goals of the system (e.g. weights, decision rules for combination, reporting).
 - The plan outlines the measures under consideration but is silent on the process that will be followed to determine the specifics.

- Montana should try to balance the technical quality of the metric to differentiate with the ease at which stakeholders can interpret the metric.
- The state might also consider supplementing the differentiation indicators with others that are used in a lower stakes reporting. This can honor some of the stakeholders' recommendations (e.g., using school climate data) without overloading the differentiation system in terms of complexity and areas of focus. Additionally, it may help educators and administrators see the intended links between lagging (i.e., outcome) data and leading indicators that may inform improvement practices.
- For those schools that are below 95% participation rate, it may worth considering differentiating targeted support schools from those schools that are targeted support for participation rate. This can help with differentiated improvement strategies that are specific to the needs of the school.
- When identifying schools for Comprehensive Support and Improvement (CSI) and Targeted Support and Improvement (TSI):
 - The state should consider how the identification process for CSI and TSI schools aligns with the annual classification/differentiation system. For example, if the system is an index, it may be tempting to identify the lowest 5% Title I schools on the lowest index scores, but that may result in a host of non-Title I schools not being identified despite their underperformance. Alternatively, CSI schools might be identified using a standalone process that is based only on Title I schools and using a subset of indicators from the accountability system.
- The CSI identification process will have implications on the TSI identification process. Because of this, the state should model the TSI identification process using data-based simulations to determine the impact of decisions.
- Develop a plan to set performance expectations for schools:
 - Will any expectations be set beyond the specific federal requirements? Will schools be given scores? What interpretations of those scores will be appropriate? How will these relate to what you are trying to accomplish for education (the theory of action).
 - Clearly defining and following a process will establish a defensible rationale for developing school-level expectations and improvement targets.

General Comments in Developing the Plan and System

We offer the following general advice to inform the state's decision-making for the system's indicators:

- Consider the level of inference with respect to the state's priorities and goals. For example, some indicators are easily collected and reported but they are low inference with respect to the attribute you want to measure (i.e., data that typically provide a count of something but do not reflect a deeper construct like engagement or climate).

- Consider the potential for corruptibility of an indicator and how the system will guard against that. For example, school climate surveys may be corruptible if safeguards are not put in place.
- Consider the ability to operationalize the measure (e.g., data collection and burden, ability to disaggregate for all subgroups (anonymized survey data, for example, cannot be disaggregated)).
- Consider the technical characteristics of each indicator. Can each indicator be measured in a reliable and valid manner? Can adjustments be made to the measures to increase or improve the reliability and validity?

We offer the following general advice in **selecting the specific indicators**:

- Growth: there is no single correct approach to growth or method that stands-out as the ‘gold-standard.’ The decision regarding which analytic approach to adopt should be informed by many characteristics, many of which are discussed in the following paper by the co-author of these recommendations:
http://www.nciea.org/publication_PDFs/CCSSO_Growth_Resource.pdf. Montana should consider at least the following:
 - Ability to support the technical properties and assumptions underlying of the growth model
 - The context and purpose for measuring growth
 - The desired model characteristics (e.g., result in a measure on the reportable scale, tolerant of missing data, resilient).
 - Ease/ clarity of interpretation
 - Ease of implementation
 - Some specific technical attributes that should be taken into consideration:
 - Good model fit and minimal error of growth estimates
 - Low correlation with prior year status
 - Stability of outcomes year to year
 - Demonstration of relatively similar distributions of growth scores for factors that should not be strongly tied to growth (e.g. ED, SWD, school size etc.)
 - Minimal floor/ ceiling effects (which are often more strongly associated with the test than the growth estimate)
- EL progress: We have seen states employ various models, but think those that incorporate growth to standard are very promising. It’s worth emphasizing that linear growth is likely not realistic (expecting the same group for a newly arrived lower performing student as for a nearly exited student).
- School/ quality student success (SQSS): We are seeing many states select credit accumulation as an option and we agree this is a promising alternative. There are also a number of excellent documents that have been written on selection/ validation of SQSS in recent months. Our colleagues Erika hall authored a paper available here:
http://www.nciea.org/publication_PDFs/CCSSO_SQSS_Brief.pdf which we think could

be helpful.

- Career ready indicators: We think promising alternatives that many states are exploring include:
 - Assessments of college/ career readiness
 - Course credit accumulation
 - Completion of pathway especially with and/or industry or 3rd party certification
 - Co-curricular learning experiences
 - Post- secondary accomplishments

We offer the following general advice with regard to **weighting**:

- Consider the intuitiveness of the system. The accountability system should be explainable, transparent, and the aggregation methods should align with the measures. There is a need to balance the ability of system to adequately and validly (based on your intended outcomes of the system) differentiate schools and incentivize behaviors based on a clear understanding of the indicators and reporting.
- Typically, this would be a judgment based process but it should be evaluated with impact data.
- We recommend the following process to help define expectations:
 - Define what an “ideal” school looks like.
 - Identify a sample of schools that should in theory be high on the “ideal” continuum, a sample of schools that have some but not all of the highly valued characteristics, and a sample of schools that are struggling.
 - Calculate scores for each of the schools
 - Explore different weighting models
 - Select the weighting model that
 - Corresponds to scores most aligned with the stated goals.
 - Adheres to the law
 - Emphasizes the indicators that are most important to the state
 - Is not biased against certain school profiles.
 - For example, one would want to see if schools that serve high percentages of students in poverty have access to favorable outcomes. If not, that may signal the model is not well specified. Note: we don’t suggest the distribution of performance is equivalent for high/low poverty schools, only that there should be some proof of concept that schools with various demographic profiles have access to favorable outcomes.

We offer the following general advice in with regard to business rules:

- Performance expectations and business rules can have a massive impact on model performance. Therefore, conclusions about model efficacy should take this into account.
- Inclusion of new students: Determining if/how newly arrived students are included in calculations is typically a policy decision. Some states have run impact data to inform their decision. (In our experience a single approach is not always more or less ‘favorable’.)

Conclusion

The above comments reflect our thoughts after a preliminary review of your draft. We understand that plan remains in development and is regularly evolving. We would be happy to work with you as you think through our recommendations and as additional questions arise. Please let us know if there are additional aspects or areas of advice that you think would prove helpful.