



BACKGROUND

During the fall of 2016, Colorado school and district performance framework reports included mean scale scores for the first time for Colorado Measures of Academic Success (CMAS) English Language Arts and Math reporting. This metric replaced using the percentage of students achieving benchmark on state summative assessments. For example, a mean score, such as 750 is now used on the frameworks as opposed to a percentage, such as 30 percent, of students meeting or exceeding expectations for CMAS. The mean scale score, while a different metric, serves to provide similar performance inferences for accountability determinations as the percent of students at benchmark.

The use of mean scale scores within the Colorado educational accountability system has been adopted based on CDE inquiry and extensive stakeholder feedback. The feedback came from the Accountability Work Group (AWG), a group of Colorado educational stakeholders, which was first convened during 2015 to provide input regarding redesign work related to the district and school performance frameworks. During the fall of 2016, the AWG was repurposed to serve as the accountability spoke committee and provided recommendations to the Every Student Succeeds Act (ESSA) process. Again, the group's recommendation was to include mean scale scores within the state plan.

RATIONALE FOR MEAN SCALE SCORES

The decision to move to use mean scale scores instead of the percentage of students at benchmark was based on stakeholder recommendations. The rationale leading to the adoption of the mean scale score included:

- The mean scale score aligns all achievement reporting within the performance frameworks to a common metric. In effect, CMAS reporting will now coincide with the use of the mean composite score that has been used for ACT reporting and will be used for PSAT/SAT reporting in the future.
- The mean scale score provides a more nuanced view of group performance. This results from the mean score being associated with distribution measures that allow for a more clearly identified group performance (i.e. relative to benchmark). Similarly, it allows for the aggregation of data across all grade levels to help further enhance the understanding of group performance.
- The mean scale score helps provide greater transparency in reporting while protecting student privacy.
- This methodology offers other advantages over percent of students at benchmark^{1,2}. Most importantly, it ensures the performance of all students is reflected in accountability metrics, not just students who perform close to the proficiency cut-scores.

It has been argued that reporting performance in terms of the percentage of students at benchmark also has a number of limitations. In a [letter](#) to Secretary John King¹, Dr. Morgan Polikoff pointed out that percent proficient is limited because it a) may lead schools to focus on students at the proficiency cut point, b) may encourage teachers to focus on moving students to minimum achievement levels, c) is not a reliable measure, d) fails to adequately measure performance gaps between subgroups, and e) penalizes schools serving more low-achieving students.

Synopsis

- Prior to 2016, the percentage of students meeting or exceeding grade level expectations was utilized within performance frameworks for accountability determinations.
- For the 2016 accountability determinations, the percent of students at benchmark was replaced by the use of mean scale scores.
- The use of mean scale scores supports data privacy, provides a more nuanced description of student performance, and aligns the various achievement reporting metrics within the performance frameworks.
- Future work by CDE will include professional development for education stakeholders to facilitate the use of mean scale scores within both accountability and improvement planning efforts.

USING MEAN SCALE SCORES FOR ACCOUNTABILITY DETERMINATIONS

The use of mean scale score for accountability determinations provides us with a method to include performance of students who are performing at the low end of the performance scale and currently nowhere near benchmark, as well as students who are doing better but can reach even higher levels of achievement. Within the frameworks, for each included assessment, the mean score (average) is calculated and reported for all cases in which the group meets a minimum n-size requirement of 16. The corresponding number of points assigned for the obtained mean score is based on the statewide school-level distribution of results that identify the 15th, 50th, and 85th percentiles for each assessment within the frameworks. A sample annotated report is available along with a scoring guide at: <http://www.cde.state.co.us/accountability/performanceframeworksresources>.

USING MEAN SCALE SCORES FOR IMPROVEMENT PLANNING

Mean scale scores may have an impact on school and district improvement planning. Specifically, trend analysis and target setting are impacted by the difference in interpretation when using mean scale scores and score distributions. It should be recognized that the use of multiple measures, including the results of local measures should always be considered during improvement planning.

When using mean scores for improvement planning, it's important that the associated score distributions are considered. Specifically, the distribution may indicate a need to focus on tier one, best-first instruction versus the allocation of resources for a more differentiated approach. For example, the performance reflected in Figure 1 indicates a larger proportion of students meeting and exceeding expectations (i.e. than those reflected in Figure 2). In contrast, the distribution of scores in Figure 2 is much narrower with most students near the meet expectation benchmark. Additional analysis, using multiple measures, which examine the percent of students performing in each proficiency level will provide more information about student needs so action steps can be appropriately thought through and developed.

Figure 1

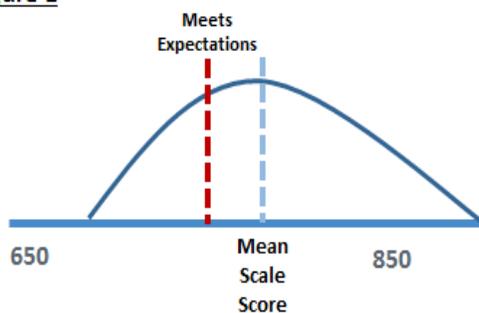
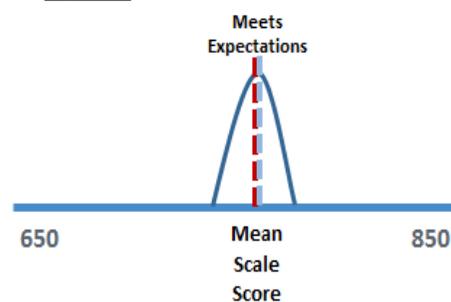


Figure 2



For more detailed information concerning the use of performance framework and other data for improvement planning please see: http://www.cde.state.co.us/uip/uip_trainingandsupport_resources.

¹Polikoff, 2016; <https://morganpolikoff.com/2016/07/12/a-letter-to-the-u-s-department-of-education>

²Ho, A. (2008). The problem with "proficiency": limitations of statistics and policy under no child left behind. *Educational Researcher*, Aug/Sep 2008, 37,6, 351-360.

Where can I learn more?

- For additional questions concerning the use of mean scale scores for accountability determinations, please contact Dan Jorgensen, Ph.D., at Jorgensen_d@cde.state.co.us
- To request training opportunities related to the use of data in the Unified Improvement Planning process, contact Erin Loften at loften_e@cde.state.co.us
- [View all CDE fact sheets](http://www.cde.state.co.us/communications/factsheetsandfaqs): www.cde.state.co.us/communications/factsheetsandfaqs