SENSIBLE REPORTING OF SPRING 2021 STATE ASSESSMENT RESULTS

June 2021

Chris Domaleski, Juan D'Brot, Scott Marion, and Michelle Boyer
Center for Assessment

Center for Assessment

National Center for the Improvement
of Educational Assessment
Dover, New Hampshire

Prepared with support from the Council of Chief State School Officers

CCSSO
Council of Chief State School Officers

SENSIBLE REPORTING OF SPRING 2021 STATE ASSESSMENT RESULTS

WHY ARE WE CONCERNED WITH TEST REPORTING?

Disruptions caused by the pandemic have impacted most every aspect of education in 2020-2021 including state testing. As states work to develop and implement novel assessment strategies to help address these challenges, it is critical to reexamine public reporting. Simply put, a 'business as usual' approach to reporting state test results is not appropriate this year.

This brief was prepared to provide guidance to state leaders to help them review and adapt ***public assessment reports*** so they will more effectively support the appropriate interpretation and uses of test scores.

WHAT ARE PUBLIC ASSESSMENT REPORTS?

Broadly, we define public assessment reports as state or district initiatives to communicate assessment results to a wide-ranging audiences. Public reporting resources typically fall into one of two categories:

- **Student Level Reports** generally include individual student reports (ISR) in paper or digital form intended primarily for parents and educators or class/school rosters included for educators and school leaders. These reports have personally identifiable information (PII); therefore they are only distributed to authorized recipients. They may also include prior student performance and summary information at the school, district, and state levels.
- **Summary Reports**
  - **Static Summary Reports** are pre-defined reports, paper or digital, with fixed features intended for broad distribution to the public. For example, summary reports may show proficiency rates at the school and district level overall and by student group. In some cases, these reports may display trend data, such as mean scale scores over the previous three years.
  - **Dynamic Reporting Tools** refer to online resources that allow users to produce customized summary reports based on selected factors. For example, users may be able to use these tools (e.g. pivot tables) to compare performance by school or student group for selected tests within or across years.

WHAT ARE THE PRIMARY CHALLENGES TO ADDRESS?

The primary challenges associated with reporting can be broadly grouped in to three categories:

- Diminished opportunity to Learn
- Changes to the test design or administration
- Changes in tested population

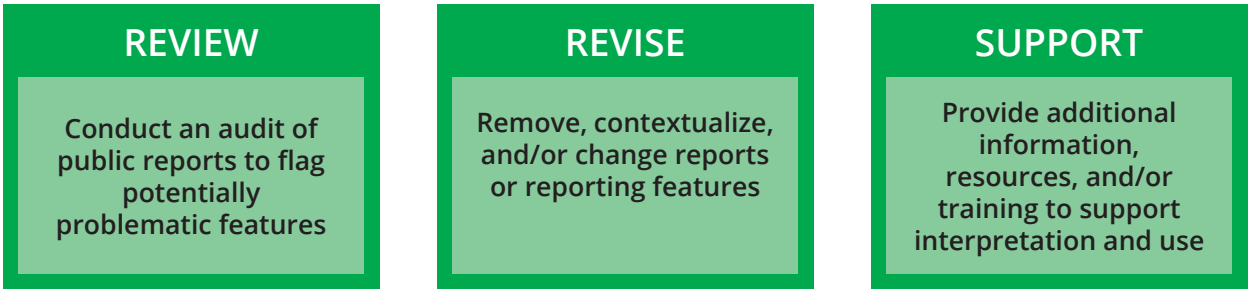
Diminished opportunity to learn refers to likely disruptions in a student's educational experience during the pandemic compared to a 'typical' year. For example, learning disruptions may have occurred due to reduced instructional time, limited access to resources to participate in remote learning, or difficulties receiving learning supports (e.g., tutoring or student services). Unfortunately, it is likely that diminished opportunity to learn will impact the most vulnerable student groups the hardest, such as students who are economically disadvantaged or English language learners.

Changes to the test design or administration refer to any modifications in the test or the any factors that affect how students interact with the test that may influence results apart from the student's knowledge and skills. For example, most testing experts agree that scores from tests administered remotely cannot be meaningfully compared with those administered in person, especially if remote proctoring is not tested and implemented. Administration may also refer to changes in the actual test, such as reducing the test length by removing performance tasks. Such changes impede the ability to compare results to prior administrations.

Finally, change in tested population refers to variance in the number and percent of students who take the state test in 2021. Most states are expecting a decrease in test participation in 2021 due to many factors including concerns that it may not be safe to test in person. The extent to which participation rates will drop in 2021 is unknown, but it is likely to be uneven and non-random across districts and schools in many states. In most years, a conventional threshold for 'full participation' is 95%. Anytime participation drops below that threshold for any level of reporting (e.g. school, district, student group), there is a potential that results are not accurate signals of student performance. The lower the participation rate, the risker it is to make inferences about performance and to compare results within year or across years. The biggest challenge with interpreting results when participation rates are considerably lower than normal involves the change in the composition of the tested population in 2021 compared to prior years.

WHAT ARE THE KEY ACTIONS LEADERS SHOULD TAKE TO ADDRESS POTENTIAL ISSUES WITH PUBLIC REPORTS?

We propose three main components in a plan to address public reporting in light of pandemic disruptions as depicted below.



Review

The review process involves evaluating all reports to determine if existing characteristics or features produce information that cannot be meaningfully interpreted in the manner intended. This may include text, tables, or graphical features. Below, we list some potential issues that may be flagged in a reporting audit.

Student Level Reports

POTENTIAL ISSUE	EXAMPLE(S)
Diminished opportunity to learn	Many states have paused student accountability due to diminished opportunity to learn, therefore references on reports to consequences such as diploma eligibility or promotion/retention should be removed.
Changes to the test design/ administration	Adjustments to the test blueprint may make interpretations of sub score or domain performance untrustworthy or no-longer relevant. Reduced test length can also increase error and reduce estimates of precision requiring.
Changes in tested population	Reports may contain explicit comparisons to summary level performance (e.g. school or district) that do not accurately reflect performance at that level due to changes in composition of test participants. Reporting percentile ranks based 2021 distributions is also an issue with changes in tested population.

Summary Reports

POTENTIAL ISSUE	EXAMPLE(S)
Diminished opportunity to learn	Many states have paused school accountability due to diminished opportunity to learn. Therefore, summary reports that contain references to accountability implications or that display results in 'accountability metrics' should be revised to avoid confusion.
Changes to the test design/ administration	Adjustments to the test blueprint may make interpretations of sub score or domain performance at the summary level irrelevant or untrustworthy. Increases in test error due to reduced test length are further compounded by reduced participation, making interpretations challenging.
Changes to the test design/ administration	Summary reports contain a mix of results from remote and in-person testing for some levels (e.g. district) which are not comparable.
Changes in tested participants	Comparisons within year between or among levels or groups are not supported due to uneven participation rates. Differences in the groups of students who are missing will affect how data are interpreted at summary levels.
Changes in tested participants	Comparisons across years between or among levels or groups are not supported due to uneven participation rates. Differences in the groups of students who are missing will affect how data are interpreted at summary levels.
Changes in tested participants (no testing in 2020)	Summary growth scores do not have the same interpretation due to implementation of alternative methodology (e.g. skip-year growth). It will be important to consider how changes in learning environment affect performance on the assessment.

Revise

Given the potential issues listed in the previous section, how can states revise public reports help ensure reports are as well-designed as possible to promote appropriate interpretation and use? We believe there are three general strategies. These are not intended to be uniform; that is, one

solution may not work well for every issue. Also, they should not be read as mutually exclusive insofar as some issues may be best addressed by drawing on multiple approaches.

- Modify Reports
- Statistically Adjust Reports
- Contextualize Reports

Modify Reports

There may be some reported elements of reports that need to be changed or removed because they are no longer relevant or have a high risk of misuse. For example, if the test blueprint no longer adequately covers a domain, removing subscores for that domain is sensible. As another example, if public reports typically display multi-year graphs of results that are no longer comparable, changing the presentation or removing the graph is a reasonable choice.

Issues about data suppression can be very complicated. First, federal or state regulations may not allow some information to be removed from reports. Second, uniform or simple 'rules of thumb' for dealing with missing information are difficult to identify. For example one school with 90% participation may be missing data from very high or low performing students, which substantially distort interpretation of the results. Another school with 70% participation may have a proportionately representative sample intact. This may prompt states to consider developing more complex criteria based on sampling theory to inform their approach. However, there are substantial drawbacks to such approaches. Not only will it be difficult to satisfy the statistical assumptions underlying these methods, it will be very challenging to implement and communicate to the public.

Statistical Adjustment

Statistical adjustments refer to a range of practices to attempt to improve the utility of public reports by modifying the information. As noted above, statistical techniques could be leveraged but in this case with the aim of producing an adjusted performance estimate for schools or groups with missing data. For example, the state could report an estimated range for the measure and unit of interest (e.g. "It is likely the school's proficiency rate is between 45% and 75%.") However, the same limitations noted above apply here as well. Accordingly, we urge great caution regarding implementing these options for public reporting. It may be more advisable to deploy these strategies for special studies than can be reported and distributed in a manner that allows the state to better communicate nuance and support appropriate use.

Contextualize the reports

Various approaches to contextualize the data may help the state improve public reports in 2021. One approach is to provide additional information. For example, the state may wish to display participation rates alongside summary proficiency rates in all instances. Another approach is to prominently add features or text to support appropriate interpretation, such as 'flags' (e.g. an asterisk or other symbol) accompanying information that should be interpreted with caution. Below are some sample statements that may be used or modified as appropriate:

Statement to accompany student level reports acknowledging potential opportunity to learn threats:

A single test score does not provide a complete or precise measure of student achievement. When interpreting results, please take into consideration other measures of student achievement. Also, consider how the conditions for learning, which may have been disrupted by the pandemic, may influence performance.

Statement to accompany presentation of summary reports in light of uneven participation:

We urge caution in interpreting summary results when participation rate are low. The standard participation threshold is 95%. When participation falls below that level, inferences about overall performance are uncertain, especially when the composition of test takers is different compared to the underlying school population in 2021 and/or compared to the test population in prior years.

Adding both contextual information and supporting text is a particularly strong approach. We recommend states work with experts and advisors, such as the state's assessment Technical Advisory Committee (TAC) to inform decisions about reporting features that fit the context for each state program and circumstances.

Support

Apart from the information available directly on public reports, education leaders may take additional actions to further promote appropriate interpretation and use.

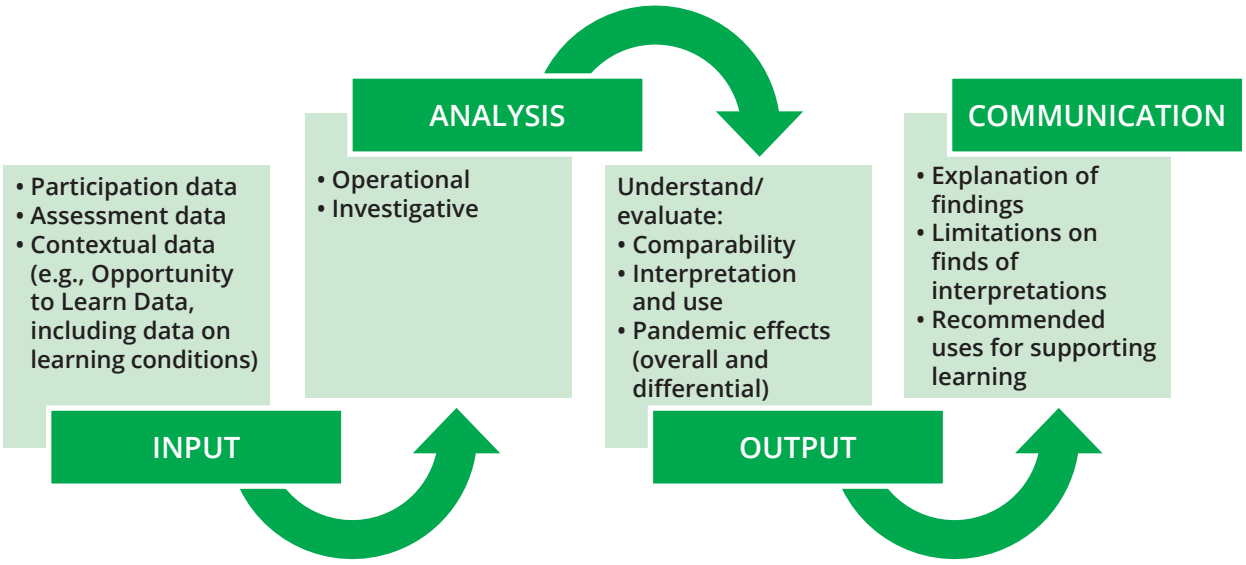
This may include the following:

- Create illustrative stories and infographics as a way to "tell stories" with the test results.
- Develop a comprehensive reporting interpretation guide, which can be referenced on each public report. Such guides often portray each report with annotations for how to best understand and use the information provided.
- Develop succinct briefs for stakeholders that include key points to guide use of public reports.
- Produce training materials, such as videos or presentations that provide guidance to report users. These materials may be consumed as stand-alone resources or repackaged as part of training programs.
- Consider producing a communications "toolkit" providing sample language, templates, and suggested documents to support district and school communications with parents.

WHERE CAN I LEARN MORE?

In this brief, we've addressed only one small part of an overall process of ensuring state test results are interpreted and used appropriately in 2021. For example, Dadey, Keng, Boyer, and Marion (2021) provide view of the 'bigger picture.' Their framework features a logic model, depicted below, that starts with efforts to explore the extent to which assessment data meet standards for technical quality and continues through a range of potential investigations that may be conducted depending on priority use case.

Analysis Framework Logic Model (Dadey et al., 2021)



This brief addresses a small but important, part of the overall framework. For example, we did not address technical issues prior to reporting, such as test scoring and scaling. Nor do we cover the full range of analyses and studies that may be informed by test scores, even though many of these studies may be disseminated publicly.

The following resources may be useful to readers interested in exploring these and other topics related to assessment in 2021.

Dadey, N., Keng, Boyer, M., & Marion, S. (2021). Making Sense of Spring 2021 Assessment Results. Dover, NH: The National Center for the Improvement of Educational Assessment.

Keng, L., Boyer, M., & Marion, S. (2020). Restart and Recovery: Assessments in Spring 2021. Council of Chief State School Officers: Washington, D.C.

Domaleski, C., Boyer, M., & Evans, C. (2020). Restart and Recovery: Guidance for Collecting, Evaluating, and Reporting Data in 2020-2021. Council of Chief State School Officers: Washington, D.C.

Betebenner, D. & Wenning, R. (2021). Understanding Pandemic Learning Loss and Learning Recovery: The Role of Student Growth and Statewide Testing. Dover, NH: The National Center for the Improvement of Educational Assessment.

Center for Assessment

National Center for the Improvement of Educational Assessment
Dover, New Hampshire

www.nciea.org