REPORT TO THE

**UTAH LEGISLATURE**

Number 2019-12

# A Performance Audit of Public Education Assessment Data

November 2019

Office of the
LEGISLATIVE AUDITOR GENERAL
State of Utah

STATE OF UTAH

# Office of the Legislative Auditor General

315 HOUSE BUILDING • PO BOX 145315 • SALT LAKE CITY, UT 84114-5315
(801) 538-1033 • FAX (801) 538-1063



KADE R. MINCHEY, CIA, CFE
AUDITOR GENERAL

November 19, 2019

TO: THE UTAH STATE LEGISLATURE

Transmitted herewith is our report, **A Performance Audit of Public Education Assessment Data** (Report #2019-12). A digest is found on the blue pages located at the front of the report. The objectives and scope of the audit are explained in Chapter I.

We will be happy to meet with appropriate legislative committees, individual legislators, and other state officials to discuss any item contained in the report in order to facilitate the implementation of the recommendations.

Sincerely,

Kade Minchey

Kade R. Minchey, CIA, CFE
Auditor General

# A Digest of a Performance Audit of Public Education Assessment Data

The Utah State Board of Education (USBE) provides an annual assessment platform and testing support to Local Education Agencies (LEA)s, which are public school districts and charter schools. USBE began providing a common assessment platform in 1999 which has changed three times in following years. The assessments measure student achievement in meeting academic proficiency standards. The resulting data is then used to assist in evaluating LEAs, schools, and teachers on their effectiveness.

Student assessment data is used and reported for various reasons and by different stakeholders. Individual student data is reported to parents and students, while broader data is available to teachers, school administrators, and Local Education Agencies (LEA)s, which are school districts or charter schools. In an LEA survey we conducted, some LEAs reported that they use the assessment data as one of multiple tools to identify schools' and teachers' performance.

## Chapter II
## Opt-Out Provisions Do Not Appear To Significantly Impact Assessment Data Accuracy

We hired independent consultants to statistically evaluate the impact of Utah's opt-out provisions on statewide standards assessment data. They found that missing scores, due to nonparticipation, have a minimal effect on data accuracy. Therefore, we currently have no evidence that suggests opt-outs are affecting the ability of policymakers to rely on the data. Nevertheless, available research and widespread anecdotal reports raise concerns that some data may be inaccurate due to disengagement, or a lack of effort by test-takers.

Recent modifications to statute allow teachers to incentivize engagement by using students' assessment scores to improve their grade for or to demonstrate competency within a relevant course. While these changes may be a promising first step to addressing concerns about disengagement, further analysis is needed. The consultants used an indirect approach to address the issue of test-taker effort and found minimal evidence of disengagement for selected groups of students. However, limitations of the available statewide testing data prevented them from determining the extent and impact of disengagement on a larger scale. Thus, further investigation is warranted.

We recommend that USBE determine the feasibility of collecting and analyzing item response data and elapsed time-per-item data from statewide standards assessments. If feasible, USBE should collect and analyze the data and consider developing and integrating its own engagement metrics into statewide standards assessments.

# Chapter III
# Assessment Data Processes
# Ensure Sufficient Validity and Reliability

USBE's development processes for questions appear effective at meeting data validity and reliability needs. USBE uses multiple review committees to examine questions at different points in question development. USBE also uses added levels of oversight to ensure data is being studied and reported appropriately and that questions are aligned with Utah's Core Standards. Finally, USBE is required to meet USDOE assessment operations and reporting requirements. USDOE's last review found that USBE either completely or substantially met all requirements except one.

REPORT TO THE

UTAH LEGISLATURE

Report No. 2019-12

# A Performance Audit of Public Education Assessment Data

November 2019


Audit Performed By:

| | |
|---|---|
| Audit Manager | Brian Dean, CIA, CFE |
| Audit Supervisor | Matt Harvey, CFE |
| Audit Staff | Michael Allred |
| | Christina Grandjacquet |

# Table of Contents

**This Page Left Blank Intentionally**

# Chapter I
# Introduction

The Utah State Board of Education (USBE) provides an annual assessment platform and testing support to Local Education Agencies (LEA)s, which are public school districts and charter schools. USBE began providing a common assessment platform in 1999 which has changed three times in following years. The assessments measure student achievement in meeting academic proficiency standards. The resulting data is then used by various users to assist in evaluating LEAs, schools, and teachers on their effectiveness.

## USBE Is Required to Conduct Statewide Annual Assessments

The Legislature has established a school accountability system that requires USBE to annually publish school performance based on performance indicators. Most of those school performance indicators come from student performance on statewide assessments. The schools annually conduct end-of-year testing to identify student proficiency in USBE defined education standards.

### Assessments Are Intended To Identify Student and School Progress

**State assessments identify student and school progress.**

Assessments are established in statute "to determine the effectiveness of school districts and schools in assisting students to master the fundamental educational skills toward which instruction is directed" (***Utah Code* 53E-4-301.5**). Assessments such as those required by statute are called summative assessments.

Typical uses for summative testing are to:

- Assist in identifying learning progress and achievement
- Assist in evaluating the effectiveness of district/school/teacher educational programs
- Assist in measuring district/school/teacher progress toward improvement goals

- Assist in making course-placement decisions

Data from summative testing is not useful for improving instruction while learning is happening because the testing happens at the end of instruction.

### Assessment Data Reporting Is Required To Receive Federal Funding

**USBE must comply with federal performance reporting requirements**

To receive federal funding, the state is required to report on school performance. Included in the reporting requirements are

- Achievement (proficiency) for all students (disaggregated by subgroups),
- Number and percent of English learners achieving English proficiency,
- Percent of students assessed and not assessed,
- Number and percent of students with significant disabilities taking an alternative assessment, and
- Names of schools identified for support and improvement.

Not included in performance reporting requirements is student specific data. When student data is shown by sub groups within a school, the sub groups must be large enough to prevent identifying any student's specific performance. If the sub group is smaller, that level of data should not be reported.

### Assessments Are Required for Students in Grades 3 through 11

***Utah Code* 53E-4-303** requires most grades to be tested in English language arts (English), mathematics (math), and science. Figure 1.1 shows current state testing requirements and the assessments.

**Figure 1.1 The State Requires Testing for Students in Grades 3-11.** Most grades are tested for proficiency in English, math, and science, with an additional writing assessment in grades 5 and 8.

| 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th | 11th |
|---|---|---|---|---|---|---|---|---|
| RISE<br>English*<br>Math | RISE<br>English<br>Math<br>Science | RISE<br>English<br>Math<br>Science<br>Writing | RISE<br>English<br>Math<br>Science | RISE<br>English<br>Math<br>Science | RISE<br>English<br>Math<br>Science<br>Writing | Aspire +<br>English<br>Math<br>Science | Aspire +<br>English<br>Math<br>Science | ACT<br>English<br>Math<br>Science |

*Source: Auditor generated based on **Utah Code 53E-4-303 -305.***
** Typically referred to as English Language Arts, we abbreviate it as English for simplicity.*

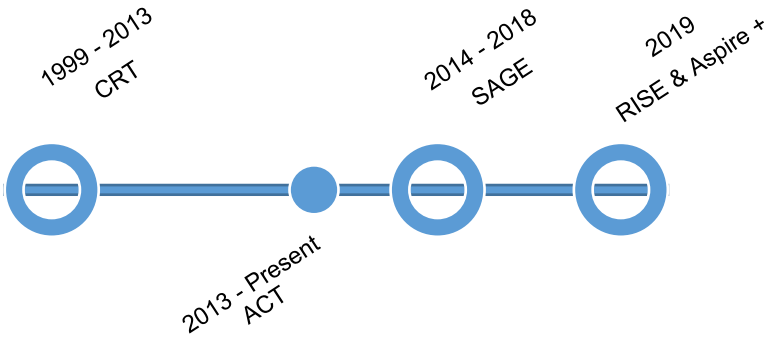
As shown in Figure 1.1, statute requires additional testing in grades 5 and 8, while grade 3 has no science requirement. Not included in the figure but required by statute is a reading assessment to be administered at the beginning, middle, and end of the school years for grades 1-3.

**Schools are required to give end-of-year student tests for grades 3-11.**

## The Assessment Platform Has Changed Three Times Since 1999

Assessment platforms have changed over time beginning with Criterion-Referenced Tests (CRT), passed into statute in 1999. The CRT tied assessment questions to the state's basic curriculum standards. Figure 1.2 shows a timeline of state testing platforms from the CRT in 1999 to Rise and Aspire Plus in 2018.

**Figure 1.2 Statewide Testing Platforms Have Changed Three Times Since 1999.** The ACT also became a requirement in 2012 for testing 11th grade. One national education assessment consultant retained by USBE told us that Utah has changed assessments less frequently than other states.




*Sources: Utah H.B. 33 (1999), Utah S.B. 175 (2013), Utah H.B. 15 (2012), USBE assessment history documents.*

**Testing platforms changed from the CRT in 1999 to SAGE in 2014 and to RISE in 2019.**

The CRT assessment required student testing in all grade levels and courses in skill areas of the state's core curriculum, while most recently, the RISE assessment was developed to replace SAGE and to test students in grades 3 through 8. The SAGE assessment was also replaced by Utah Aspire + to test students in grades 9 and 10. Finally, grade 11's college readiness student assessment requirement is met by the ACT. Utah had additional student testing requirements over the years for testing areas including writing and civics.

School accountability evaluations have also changed over time. In 2011, the Legislature passed requirements for USBE to grade schools based mostly on assessment data. In 2013, USBE produced its first school grades based on students' proficiency, proficiency growth, and graduation rates. Figure 1.3 shows the percentage weights by which schools were graded in 2013. For example in 2013, an elementary school could earn 300 points for student proficiency, 150 for proficiency growth, and 150 for proficiency growth of those below proficient. Later, school accountability indicators evolved to include additional indicators and different percentage weights.

Note in Figure 1.3 that the percentage weights for school grades do not change drastically across the years, but the changes in weights reflect a greater emphasis on students' proficiency improvement rather than only meeting proficiency. Additional school accountability indicators include English learner progress for all schools as well as ACT scores and advanced coursework for high schools.

**Figure 1.3 School Accountability Indicators Have Evolved with Time.** Indicators have evolved since 2013 to place more emphasis on student improvement.

| | Measure | 2013 | 2014 | 2018 |
|---|---|---|---|---|
| Elementary & Middle Schools | Proficiency | 50% | 50% | 37% |
| | Proficiency Growth | 25 | 25 | 37 |
| | Growth of Lowest Performing 25%* | 25 | 25 | 17 |
| | English Learner Progress | | | 9 |
| | **Total** | **100%** | **100%** | **100%** |
| | **Measure** | **2013** | **2014** | **2018** |
| High Schools | Proficiency | 40% | 33% | 25% |
| | Proficiency Growth | 20 | 17 | 25 |
| | Growth of Lowest Performing 25% | 20 | 17 | 11 |
| | ACT Performance | | 17 | 11 |
| | Graduation Rate | 20 | 17 | 11 |
| | Advanced Coursework | | | 11 |
| | English Learner Progress | | | 6 |
| | **Total** | **100%** | **100%** | **100%** |

*Source: USBE School Grade Reports*
** In years prior to 2018, the measure was focused on proficiency growth for students below proficiency levels, not just the lowest performing 25 percent.*

The change in emphasis from proficiency to proficiency growth reflects a larger shift in educational policy nationally. Proficiency essentially measures where the student is at a point in time, but does not necessarily account for the progress of students and the quality of instruction the student has received. Proficiency growth may better quantify the impact of a school on student learning for a specific period, and proficiency growth for the lowest performing 25 percent of students may encourage schools to ensure their lowest performing students are not falling behind.

**Proficiency growth helps to identify a school's success in improving student learning.**

While ***Utah Code* 53E-5-204** has required that USBE assign grades to schools based on accountability indicators, that requirement was suspended for the 2017-2018 school year. The statute allows that in a year where USBE is establishing a new baseline, USBE is not required to assign grades. The statutory requirement will resume for following years unless the Legislature suspends it again.

## Assessment Data Has Been Used Relatively Consistently

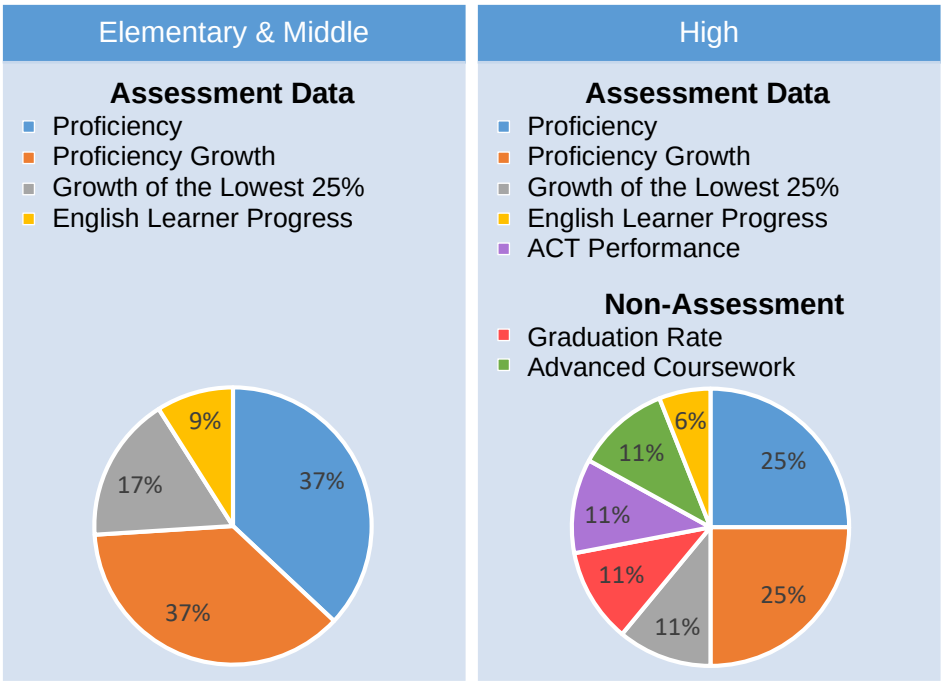
**Data is used by students, parent, teachers, school administrators and LEAs.**

Student assessment data is used and reported for various reasons and by different stakeholders. Individual student data is reported to parents and students, while broader data is available to teachers, school administrators, and Local Education Agencies (LEA)s, which are school districts or charter schools. In an LEA survey we conducted, some LEAs reported that they use the assessment data as one of multiple tools to identify schools' and teachers' performance.

Assessment data, coupled with cost and other performance data, is also being used in a tool being developed by the Office of the State Auditor (OSA). OSA is working to integrate financial, operational, and student performance information to help stakeholders better understand where the money goes in public education. OSA expects the data will ultimately provide insight into school-, classroom-, and student-level resource allocation and associated performance outcomes. Initially, OSA is focused on making dashboards available to LEA personnel but will later work with stakeholders to develop dashboards for the public and other policymakers.

Assessment data is used by USBE to determine individual school achievement levels. As shown in Figure 1.4, USBE currently grades schools based on a mix of proficiency, proficiency growth, and other indicators.

**Figure 1.4 School Accountability Indicators Mostly Consist of Student Assessment Data.** Proficiency measures students' understanding of content areas, whereas proficiency growth measures whether that understanding has improved over time.




*Source: USBE school grade data.*

As shown in Figure 1.4, proficiency growth (including growth of the lowest 25 percent students in proficiency) accounts for more of a school's grade than proficiency itself.

**Proficiency growth accounts for more than proficiency in school grades.**

School grades can help identify how schools perform compared to others and which schools may need additional resources or interventions. The lowest performing schools are targeted by three separate programs:

**State School Turnaround**: If a school is scored in the lowest three percent of state schools for two years in a row, the school is placed into turnaround status. Schools placed into turnaround status are given two years of consecutive improvement over a three-year period to exit turnaround status. During that time, the schools are allocated additional resources. If schools do not improve enough to exit turnaround status, USBE can intervene by changing management, by taking over the school, or by other measures such as providing more time or closing the school.

**Data is used to identify schools in need of additional resources for improvement.**

**Comprehensive Support and Improvement (CSI)**: Once every three years, the lowest performing 5 percent of Title I schools are identified based on schools' average performance. Those schools must work with stakeholders to identify an improvement plan which will be monitored by USBE.

**Targeted Support and Improvement (TSI)**: Specific student groups within a school that perform in the lowest 5 percent of all students for two years in a row fall in to this program. As a result, LEAs must identify an improvement plan to improve those groups' outcomes. Improved outcomes for two years in a row will allow the group to exit TSI status.

## Audit Scope and Objectives

**We were asked to identify the reliability of the data for making policy decisions.**

We were asked to identify the reliability of current assessment data for making policy decisions and to report how assessments have changed over the years. Specifically, we evaluated:

- Assessment data uses and history (Chapter 1)
- Impacts of parental exclusion on data accuracy (Chapter 2)
- Assessment reliability and accuracy (Chapter 3)

# Chapter II
# Opt-Out Provisions Do Not Appear to Significantly Impact Data Accuracy

We hired independent consultants to statistically evaluate the impact of Utah's opt-out provisions on statewide standards assessment data. They found that missing scores, due to nonparticipation, have a minimal effect on data accuracy. Therefore, we currently have no evidence that suggests opt-outs are affecting the ability of policymakers to rely on the data. Nevertheless, available research and widespread anecdotal reports raise concerns that some data may be inaccurate due to disengagement, or a lack of effort by test-takers.

Recent modifications to statute allow teachers to incentivize engagement by using students' assessment scores to improve their grade for or to demonstrate competency within a relevant course. While these changes may be a promising first step to addressing concerns about disengagement, further analysis is needed. The consultants used an indirect approach to address the issue of test-taker effort and found minimal evidence of disengagement for selected groups of students. However, limitations of the available statewide testing data prevented them from determining the extent and impact of disengagement on a larger scale. Thus, further investigation is warranted.

**We did not find evidence to support concerns about opt-outs, but more study is needed to understand potential impacts of disengagement.**

We recommend that the Utah State Board of Education (USBE) determine the feasibility of collecting and analyzing item response data and elapsed time-per-item data from statewide standards assessments. If feasible, USBE should collect and analyze the data and consider developing and integrating its own engagement metrics into statewide standards assessments.

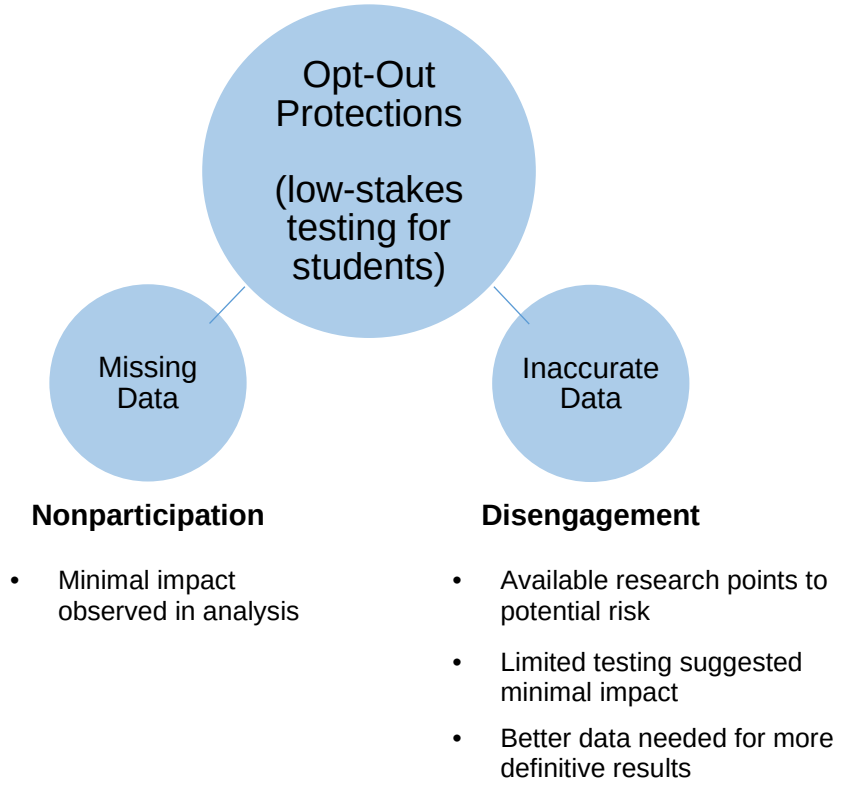
## Utah's Opt-Out Provisions Have Two Inherent Risks

To identify risks to the accuracy of statewide assessment data, we met with various public education stakeholders. Their main concerns were: (1) missing data due to parental opt-out, and (2) inaccurate data due to disengagement.

**Utah's opt-out provisions create a low-stakes testing environment by protecting students from direct consequences for nonparticipation or disengagement.**

Utah's opt-out provisions protect students from direct consequences for opting out of statewide standards assessments, and for not trying their best on statewide standards assessments. These protections create low-stakes[1] testing conditions for students and come with two inherent risks to data accuracy: (1) nonparticipation, which yields missing data; and (2) disengagement, which yields inaccurate data. Figure 2.1 summarizes these two risk areas and our related findings.

**Figure 2.1 Opt-Out Protections Have Two Inherent Risks: Missing Data and Inaccurate Data.** Utah's opt-out provision prohibits direct consequences to students for nonparticipation and test performance.




*Source: Auditor generated summary of risks associated with opt-out provisions*

**Nonparticipation leads to missing values in the assessment data.**

The first risk, missing data, is primarily a result of nonparticipation. Parental exclusion, or opt-out, is the leading cause of nonparticipation. ***Utah Code*** 53G-6-803(9), allows parents to exclude their children from participating in the assessments and states:

[1] Low stake assessments are meant to identify how well an education or school system is doing but have no direct consequences on the test-takers.

> At the request of a student's parent, an LEA shall excuse a student from taking an assessment that … is mandated by the state … or requires the use of (A) a state assessment system; or (B) software that is provided or paid for by the state.

**Test-taker disengagement can lead to data inaccuracies.**

The second risk, inaccurate data, is associated with disengagement, which available research suggests is a potential risk for low-stakes tests. Statewide standards assessments are considered low-stakes for students because ***Utah Code*** 53E-4-303(4)(a) prohibits using assessment scores to determine a student's academic grade for a course or to determine whether the student will advance to the next grade level. Although limited analysis by the consultants provides some evidence of minimal disengagement, further analysis is merited.

To understand how opt-outs and disengagement impact statewide standards assessment data, we hired a team of three university professors (consultants) to conduct statistical analyses of SAGE data and had their work peer reviewed[2] for accuracy. We tasked the consultants with evaluating how nonparticipation and disengagement affect the legitimacy of the data. Appendix A has the consultants' report.

## Impacts of Missing Data Due to Nonparticipation Appear Minimal

Statistical analyses of missing data suggested that nonparticipation has a minimal effect on data accuracy. In their report, the consultants concluded that "opt out has a minimal threat to the validity of inference based on observed test scores." Thus, we found no evidence of opt-outs affecting the ability of policymakers to use the data. Stakeholder concerns about parental opt-outs negatively impacting statewide standards assessment data focused on two primary assumptions: (1) that the data cannot be trusted to accurately reflect schools' performance because of missing values for students who were opted-out of testing, and (2) that school accountability grades are artificially low because opted out students are primarily high performers. To address these concerns, we asked the consultants to

**Missing data due to nonparticipation appears to have minimal impact on data used for accountability.**

[2] Special thanks to Randy E. Bennett for his independent review of the consultants' opt-out study.

analyze how different participation conditions would have affected assessment data accuracy.

Initially, the consultants focused only on missing scores for students who were opted-out of SAGE testing. However, we suggested they analyze missing scores for nonparticipation generally after they found no significant difference between students who were opted out and students who did not take the test for other reasons.

The consultants further determined that although nonparticipation rates vary across geographic areas within the state, nonparticipation is predictable and significantly correlated with four demographic variables (student gender, socioeconomic status, ethnicity, and mobility), school type, and grade level. These findings allowed them to estimate the significance of missing scores in the data due to nonparticipation.

**Replacing Missing Scores with Plausible Estimates Had No Significant Impact.**[3] Consultants used accepted statistical methods of multiple imputation to evaluate the effect of missing scores on the data. They built a hypothetical data set using substituted scores for nonparticipants and compared it to the original data set with missing scores. The scores were estimated using six statistically significant predictive variables. This allowed them to see how the missing scores might affect school rankings, particularly for schools at risk of falling into the bottom fifth percentile of schools. Using two different methods, they compared actual school rankings (based on data with missing scores) to projected school rankings (based on data with estimated scores). They found the rankings were highly correlated and that the missing data had a minimal effect on school classifications. Based on these findings, the consultants concluded that:

**Consultants found the missing data had little effect on how schools were classified for school improvement.**

> Overall, opting out had a minimal effect in terms of decisions about which schools would be identified for Comprehensive Support and Improvement. Because a school has to be below the cut-off for three consecutive years before being classified in this category, the risk of false-positive decisions is extremely small.

The consultants' conclusions are based on current Comprehensive Support and Improvement policies. Consequently, there may be larger

[3] See pages [43 – 48] in Appendix A for more detail.

policy considerations that this audit does not address. Additional statistical analysis may be needed should the Legislature choose to change or strengthen those policies.

**Students Opting Out Tended on Average to Score Lower than Peers in Previous Tests.**[4] Consultants tested for differences between the SAGE scores of seventh graders who would eventually opt-out of testing as tenth graders, and the SAGE scores of their seventh-grade peers. The consultants found that the seventh graders—who would eventually opt out of SAGE testing as tenth graders—had significantly lower SAGE scores than their peers, meaning they were lower performing students. This test contradicts some anecdotal reports we heard that primarily high performing students are opted out of assessments. In their conclusions, the consultants stated:

> We did not find evidence to support the anecdotal claim that higher-achieving students opt out more frequently than their lower-achieving counterparts. These results are contrary to the anecdotal claims that higher achieving students have a greater tendency to opt out than lower achieving students.

**Consultants found no evidence to support the claim that high performing students opt out more than their peers.**

## Further Investigation is Needed to Determine The Prevalence and Impact of Disengagement

Closely related to concerns of students opting out of assessments are concerns about test-taker disengagement, sometimes referred to as "mental opt-out." One of the primary purposes of statewide standards assessments is to measure student knowledge, but when students disengage from the tests, their scores may not be reflective of their knowledge. Regarding this issue, the consultants reported:

> At present, we do not know to what extent the lack of test-taking motivation poses a threat to the validity of inferences about school effectiveness that are likely to be made by policy makers and other stakeholders based on scores obtained from state-administered tests in Utah…If it is widespread, then this lack of motivation is likely to be a greater validity threat than nonparticipation due to opt

**Disengagement may pose a threat to data validity; however, the level and extent of disengagement are unknown.**

[4] See pages [48 – 50] in Appendix A for more detail.

out…We believe that this is a critical issue which warrants further investigation.

**Teachers have the option to incentivize test-taker engagement after statutory amendments by Legislature.**

Legislators passed House Bill 118 during the 2019 Legislative General Session in an effort to encourage effort by students who participate in statewide standards assessments. The bill modified ***Utah Code* 53E-4-3** to allow teachers to incentivize student engagement in the tests by using students' scores to improve their grade for or demonstrate competency within a relevant course. Such a policy may serve to reduce student disengagement, but better data is needed determine its impact.

We surveyed LEA assessment directors and they estimated that around one quarter of test takers deliberately underperform on state tests. Moreover, available research identifies test-taker disengagement as a threat to data accuracy. Although the consultants' indirect statistical analyses of student effort found no evidence of significant disengagement, these results are inconclusive. Due to limitations of the available data, further analysis is impossible. Consequently, USBE should begin collecting additional data so further analysis of disengagement can be conducted to determine its frequency and impact.

### Stakeholders Are Concerned About Test-Taker Disengagement And its Effects on Data Validity

**Some public education stakeholders were concerned about the risk of inaccurate data resulting from perceived test-taker disengagement.**

Some stakeholders we interviewed reported they lack confidence in statewide assessment data because perceived disengagement may yield inaccurate data, which could underrepresent student knowledge and negatively bias the accuracy of the data. To better understand this perception, we surveyed assessment directors from school districts and charter schools statewide and asked them to estimate rates of test-taker disengagement. Specifically, we asked them to estimate the percentage of students in their respective LEAs who deliberately underperform while taking a state test. Assessment directors from 87 charter schools and 40 school districts provided estimates, which are summarized in Figure 2.2.

**Figure 2.2 Assessment Directors Estimates of Disengagement.** Assessment directors estimated that roughly a fifth to a quarter of students deliberately underperform on statewide assessments.

| Charter School Assessment Directors Estimated That | |
|---|---|
| 21% | of students deliberately underperform while taking a state test. |

| School District Assessment Directors Estimated That | |
|---|---|
| 28% | of students deliberately underperform while taking a state test. |

*Source: LEA Assessment Director Survey from our office*

School districts had higher overall estimates of disengagement, but averages between LEA types only differed by 7 percent. The median percentages (15 percent for charter schools and 25 percent for school districts) indicated a broader range and distribution of the reported estimations, and a need for additional review. To further understand this issue, we asked the consultants to look for evidence of disengagement in their study.

### Limited Statistical Analyses of SAGE and LEA Data Did Not Show Significant Evidence of Disengagement[5]

To assess the impacts of disengagement, the consultants analyzed the correlation between students' scores on mandatory (high stakes) district or charter school summative assessments and their scores on statewide SAGE assessments (low stakes). They concluded that students' SAGE scores were highly correlated with their district or charter school scores, meaning there was no statistically significant evidence of disengagement or deliberate underperformance by the students on the statewide standards assessments. Nevertheless, their ability to draw conclusions were limited by the available data and they recommended additional data collection and investigation of disengagement.

**Consultants found high correlation between student scores on SAGE and required LEA-level assessments.**

[5] See pages [50 – 54] in Appendix A for more detail.

To conduct their disengagement analysis, consultants sampled two school districts and two charter schools. The consultants selected the districts and charters based on their use of additional assessments for end-of-year testing. Those districts and schools reported using their own end-of-year assessments in addition to administering the statewide assessments.

**Some LEAs use additional assessments to supplement statewide assessment results.**

LEA staff cited opt-out as one reason they use their own assessment software. One LEA assessment director said specifically that because their opt-out rates were high, the LEA needed an alternative test to measure student progress. LEAs can get more complete data with their own assessments because the statutory opt-out provision applies only to state funded and/or required tests.

The consultants used the selected district and charter school data to determine how well performance in LEA-required testing matched with performance in SAGE testing. The consultants found significant correlation between LEA assessment performance and SAGE performance.

Correlation measures the extent to which variables tend to change together. In this case, a strong correlation of student performance between district and state testing would be evidence that the tests are measuring similar things. A strong correlation would also indicate that student performance does not drastically change between the high stakes district and charter school tests and the low stakes state test. Based on the limited data examined, the consultants concluded that there was no significant evidence of disengagement in the SAGE test results.

### Disengagement Metrics Could Help USBE Understand Impacts on Assessment Data

**USBE can address disengagement concerns by collecting and analyzing additional testing data.**

Given existing concerns about test-taker disengagement and its effects on the validity of test score interpretations, USBE should consider working with testing platform providers to add engagement metrics to statewide assessments. Specifically, they should begin collecting item response data and elapsed time-per-item data from assessments for evaluation. The consultants suggest this data should be easy to collect. They also reviewed some available research on measuring and evaluating test-taker disengagement and found multiple methods available for evaluating the prevalence of disengagement.

In a limited review, we found one international provider of student assessments with two engagement metrics in its student assessments. The implemented metrics measure: (1) percent of disengaged responses, and (2) estimated impact of disengagement. The provider also issued guidance on using the metrics to increase data credibility. Recommendations ranged from retesting to invalidating test results depending on the magnitude of disengagement.

**We found one testing vendor that uses disengagement metrics to inform assessment results.**

Based on available literature, recommendations by the consultants, industry practices, and widespread concerns about disengagement we recommend that USBE consider collecting and analyzing data to evaluate its extent and impact. The following recommendations are intended to help USBE evaluate the effects of disengagement on the accuracy of statewide standards assessment data.

## Recommendations

1. We recommend that the Utah State Board of Education determine the feasibility of collecting and analyzing item response data and elapsed time-per-item data from statewide standards assessments.

2. We recommend that if the Utah State Board Education determines collecting item response and elapsed time-per-item data is feasible, it should: (1) collect the data for analysis, (2) use the data to evaluate the extent and impact of test-taker disengagement on statewide standards assessments, and (3) consider developing and integrating its own engagement metrics into statewide standards assessments.

3. We recommend that if the Utah Legislature is concerned about the potential impacts of disengagement on data validity, that it consider requiring the Utah State Board of Education implement the above recommendations.

**This Page Left Blank Intentionally**

# Chapter III
# Assessment Data Processes Ensure Sufficient Validity and Reliability

The Utah State Board of Education's (USBE's) development processes for questions appear effective at meeting data validity and reliability needs. USBE uses multiple review committees to examine questions at different points in question development. USBE also uses added levels of oversight to ensure data is being studied and reported appropriately and that questions are aligned with Utah's Core Standards. Finally, USBE is required to meet the U.S. Department of Education (USDOE) assessment operations and reporting requirements. USDOE's last review found that USBE either completely or substantially met all requirements except one.

In addition to reviewing USBE assessment processes and oversight, we met with and surveyed school administrators to discuss concerns they had with assessment data. One common theme we heard was that assessment data is inaccurate because of parents' ability to opt children out of state tests, which we discuss Chapter II.

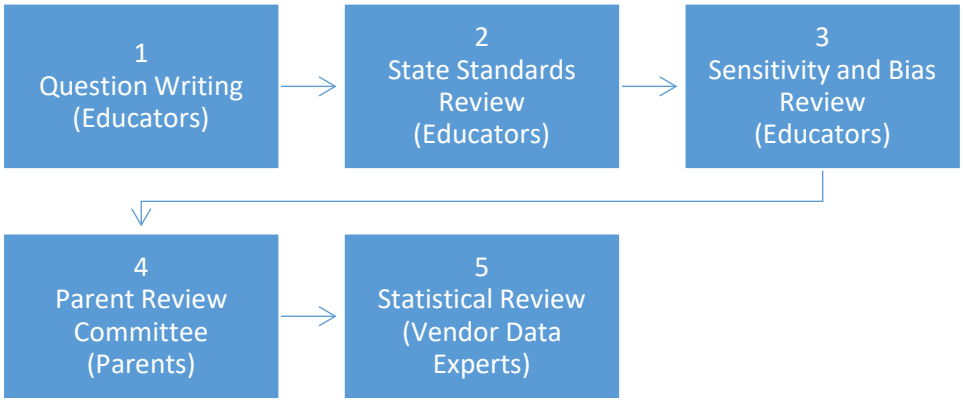
**We both met with and surveyed school administrators.**

## Test Questions Undergo Extensive Review

Assessments rely on the quality of the questions used. For an assessment to be valid, the questions must be designed to test the desired content. Assessment validity is the degree to which questions measure what they are intended to measure, whereas assessment reliability is the consistency of test scoring and performance levels. To meet validity and reliability objectives, USBE test questions go through development processes from the writing of the question to statistical review of student responses to the question. The general process used by USBE is shown in Figure 3.1.

**Every question is reviewed by educators, parents, and data experts.**

**Figure 3.1 Questions Are Written by Educators and Reviewed by Educators, Parents, and Data Experts.** See the corresponding paragraphs below for further detail on each step.




*Source: Auditor Generated*

1. The first major step in developing assessment questions begins with Utah educators gathering in workshops to write the questions. Educators receive training on question writing and Utah's Core Standards. From there, educators begin writing questions targeted at needed Core Standard areas.

2. Next, another set of educators review the assessment questions to identify the standard being tested and the question's difficulty level. This is to ensure that the assessment can reasonably determine the student's depth of knowledge concerning a specific standard.

3. Educators then review the questions for bias and sensitivity issues. Reviewers are trained to look for stereotypes, sensitive or controversial subjects, promotion of specific lifestyles or behavior, promotion of dangerous activities, ethnocentrism, specialized concepts or terminology unrelated to the standard, and language accessibility. The reviewers are also selected to ensure a variety of backgrounds on the committees.

4. Once the questions have been reviewed for standard alignment and bias, a parent review committee examines the questions. ***Utah Code*** 53E-4-303(5) requires that "The state board shall establish a committee… to review all standards assessment questions." The committee consists of 15 Utah public education parents selected by

the USBE chair, the Legislative House Speaker, and Legislative Senate President.

5. Finally, the questions are field tested to gather information on how the questions perform in a testing environment. The testing vendors use the resulting data to perform statistical analyses on each question. The analyses ensure that the questions are scored by the software as intended, that the questions meet the established difficulty level, and that the questions do not discriminate between student subgroups.

## USBE Engages Additional Accountability and Assessment System Review

**USBE uses advisory committees and independent evaluations to provide assessment guidance.**

In addition to USBE question development oversight, USBE uses two key advisory committees to guide the assessment process. USBE has also contracted for independent evaluations of its assessment questions.

The Assessment and Accountability Policy Advisory Committee advises USBE in current assessment and accountability policies and rules as well as the rollout of related changes and requirements. The committee is made up of education stakeholders such as school and district administrators, board members, assessment directors, an educator, and a parent.

USBE also uses a Technical Advisory Committee (TAC) to advise it in meeting technical requirements related to assessment and accountability data accuracy. The TAC contains experts in education data analysis. The TAC reviews the question analyses conducted by the assessment software provider to ensure the analyses are appropriate and methodologically sound.

USBE also contracted with evaluators in 2016 for an independent study of the degree to which Utah's assessment tests Utah Core Standards. As stated earlier in the chapter, USBE uses alignment committees to ensure that questions are aligned with standards, but the evaluators provided additional assurance that the questions and the assessment itself function appropriately. Evaluators found that the assessment in most cases adequately tested the standards intended.

In cases where study evaluators found the assessment lacking, they found that the standards could benefit from increased question coverage at high difficulty levels and that some questions were used too often (which means they have a shorter usable life). USBE responded by contracting for an assessment which would allow more control over the assessment algorithms and require fewer overall questions.

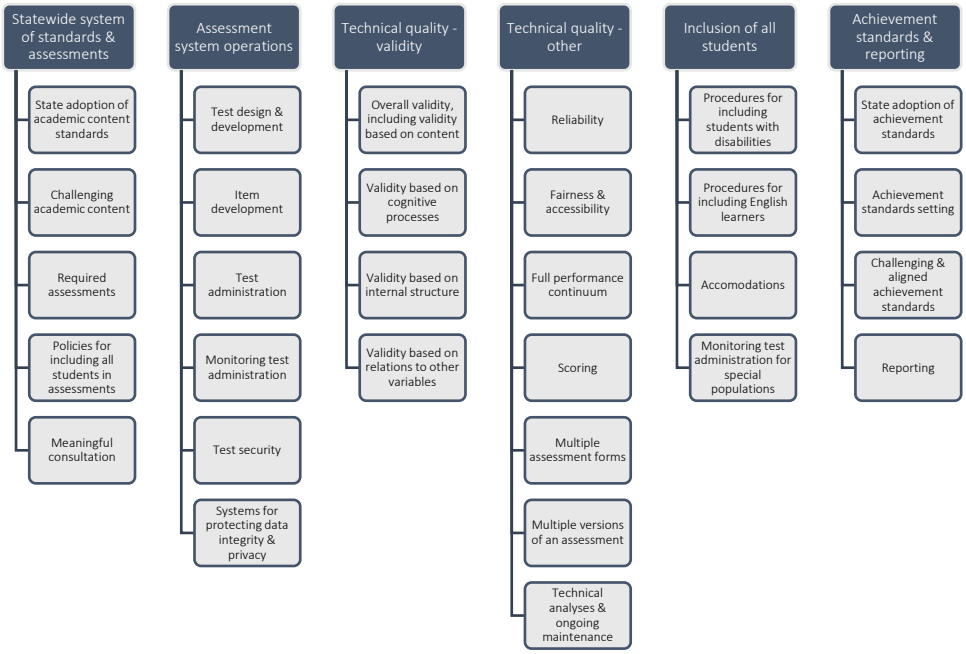
## USDOE Requires State Accountability Systems to Be Peer Reviewed

**Accountability system peer reviews are required by the U.S. Department of Education.**

The U.S. Department of Education (USDOE) also requires each state to submit to a peer review of its statewide accountability system. USDOE selects education professionals to review state submissions and to advise USDOE officials on state compliance to USDOE accountability requirements. Based in part on peer input, USDOE makes final determinations on the adequacy of the state's accountability processes.

USDOE's reporting requirements appear extensive. They require USBE to address assessment elements such as education standard adoption, system operations, student inclusion, achievement standards and reporting, and assessment validity, reliability, fairness and scoring. Figure 3.2 shows USDOE requirement elements.

**Figure 3.2 U.S. Department of Education State Assessment Requirement Elements.** The USDOE requires states to submit evidence demonstrating compliance in each of the elements listed.




*Source: A State's Guide to the U.S. Department of Education's Assessment Peer Review Process. p. 29. 2018. (Adapted for clarity and brevity)*

In USBE's most recent peer review, USDOE found that USBE had met all requirements except for one main unresolved concern which USDOE said only partially met requirements. USDOE's concern was that science was not being tested consistently in high school due to USBE not requiring specific science classes tied to specific grades (and therefore not testing specific science standards for all students in those grades). For example, not all 9th grade students are required to take the same science course and therefore do not receive the same science test. USBE has since changed its 9th and 10th grade assessments which now test the same science standards for all students in high school.

**USBE fully or substantially met all USDOE requirements except one. USBE has since changed its assessment to meet the requirement.**

USDOE also had a concern that USBE substantially met question bank requirements but did not fully demonstrate that its question bank was robust enough to meet all content standards. USBE has been addressing the concern by targeting question development for underdeveloped content areas.

**This Page Left Blank Intentionally**

# Appendices

**This Page Left Blank Intentionally**

# Appendix A
# Consultant Report

**This Page Left Blank Intentionally**

# The Incidence and Impact of Missing Data on Utah's 2018 SAGE Tests

# Due to Parental Opt-Out and Self Opt-Out

A report presented to the

Office of the Legislative Auditor General

by

Lane Fischer

Ross A.A. Larsen

Richard R Sudweeks

October 24, 2019

## TABLE OF CONTENTS

## BACKGROUND AND RATIONALE

In compliance with Utah law and the federal Every Student Succeeds Act (ESSA), the Utah State Board of Education administers a series of grade-level tests to students in the public schools near the end of each school year. These tests are part of the accountability system implemented by the state to (a) rate the effectiveness of schools which receive public funding, (b) identify schools that are underperforming, and (c) inform policymakers and interested patrons about the performance of the various schools.

Utah law includes a statute [53G-6-803(9)(a)] which specifies that parents of students enrolled in grades 3-11 in public elementary and secondary schools may choose to have their child exempted from completing one or more of the state tests. To exercise this option, a parent must submit a written request to their child's school in which they name the child and specify one or more of the tests from which they want the child to be exempted. When parents exercise this option, the score for each untested child becomes a missing value in the resulting data file and may affect the generalizability of the test results. For instance, missing scores have the potential to lead to biased conclusions about the average level of student achievement in the public schools across the state. Missing scores may also impact decisions about which schools are underperforming.

The degree to which the missing scores limit the generalizability of the test results will likely depend upon (a) what proportion of the scores are missing, (b) whether missingness is predictable based on other variables in the data set, and (c) whether the missing scores are related to the very variables that are measured by the SAGE tests. There are no established guidelines in the statistical literature about how much missingness is too much. From a statistician's point of view, the type of missingness is more important than the quantity. Although there are no established guidelines specifying the point at which the proportion of missing data becomes

problematic, there are widely accepted principles regarding the type of missingness and how the missingness is related to other variables in the data set. These guiding principles resulted from pioneering work by Donald Rubin more than 40 years ago and subsequent refinements made by Rubin and his associates (Rubin, 1976; Little & Rubin, 1987, 2002, 2020).

According to Rubin and his colleagues, there are three main types of missing data:

1. If the missing scores are randomly scattered throughout the data in an unpredictable manner, then they are considered *ignorable* because they are unlikely to limit the generalizability of the obtained scores (Little & Rubin, 2019).
2. If the missing scores tend to follow some systematic pattern that can be predicted from other variables in the data set, then the missingness is *ignorable, but predictable*.
3. However, if the missing scores are related to the very variables that are measured by the SAGE tests, then the missingness is *nonignorable* and users should be cautious about drawing conclusions based on the obtained scores. For example, if students who are above average on the trait measured by the test tend be systematically missing from the test file, then the mean of the obtained scores will be an underestimate of average achievement in the schools. Conversely, if students who are below average on the trait measured by the test tend to be systematically missing from the test file, then the mean of the obtained scores will overestimate the actual level of achievement.

Since the results of these annually administered accountability tests are publicly available and are part of the evidence used to rate the effectiveness of individual schools, the tests have serious potential consequences for the schools. However, the tests do not have any direct personal consequences for the students who complete them. In fact, statute 53E-4-303(4)(a)

prohibits the use of scores from these state tests in determining a student's grade in a course or in making decisions about whether a student should or should not be advanced to a higher grade level. Therefore, these end-of-year assessments are high-stakes tests from the point of view of teachers and school administrators, but they are low-stakes tests from the students' point of view.

Because of the perceived low-stakes nature of the tests, some students may not feel motivated to expend effort to answering the questions or problems posed. Unless a student exerts sustained effort, their test performance is likely to result in a score which underrepresents what they know and can do. This lack of effort not only leads to negatively biased estimates of students' level of proficiency, it also poses a threat to the validity of inferences about the performance of individual schools (Wise & DeMars, 2005; 2010).

## STATEMENT OF PURPOSE

The main purpose of this study was to examine the effects of the provision in Utah law [53G-6-803(9)(a)] which allows parents to formally request that one or more of their children be exempted from taking the state-administered tests. A secondary purpose was to assess the impact of unmotivated performance on the part of students who did take the tests but did not give their best efforts in completing them.

### Research Questions

The research reported in this study focused on answering five questions about missingness due to nonparticipation in the 2018 administration of the SAGE tests:

1. What proportion of the students in Utah public schools who were eligible to complete the SAGE tests administered by the state in the spring of 2018 did not participate in the testing?

2. How were the missing scores distributed across various subgroups including the different types of schools, different grade levels within the schools, and other demographic groups, and what other variables were correlated with the missingness?
3. How did nonparticipation impact decisions about which schools should be categorized as Comprehensive Support and Improvement (CSI) schools?
4. What evidence exists that either supports or does not support the anecdotal claim that the nonparticipation rate is larger among high achieving students than among low achieving students?
5. What evidence is there, if any, that students lack motivation to expend their best effort in responding to the SAGE tests because of their perception of the low stakes nature of the test?

**Available Data**

To conduct this study, we were given access to test score files which included data for 394,915 Utah students in grades 3-11 in the spring of 2018. The files were previously deidentified and did not include students' names. Students enrolled in private schools were not required to take the state tests and therefore were not included either in the files or in the analysis. The files did not include any students in kindergarten or in first or second grades since SAGE tests are not administered in those early grades. Since local education agencies were given the option to decide whether or not to test students enrolled in the eleventh grade, some eleventh graders took the tests and their scores were included in the file which we analyzed. However, a large majority of eleventh graders did not participate in the 2018 SAGE testing. We have included the rate of nonparticipation of eleventh graders in the descriptive statistics that we report in this document, but none of the eleventh graders were included in any further statistical analyses.

Although twelfth graders are not required to take the SAGE tests, the files did include scores for some twelfth grade students who were enrolled in classes consisting mainly of younger students such as twelfth graders taking an eleventh-grade math or English class. The twelfth graders who were included in the data set were excluded from all of our analyses based on rationale that they are not required to take SAGE tests and therefore may not give their best effort in responding to the test. Furthermore, they represented only a small fraction of the total number of the students in their grade. Therefore, the population of interest in this study was defined as all students enrolled in grades 3-10 in Utah public schools (including traditional district schools, charter schools, and virtual schools) when the tests were administered in the spring of 2018 except in one analysis that used the 2015 test scores. By state law, charter schools and traditional district schools are both classified as public schools. Consequently, we have used the label *district schools* in this report to refer to traditional, non-private, neighborhood schools that have a defined catchment area and are not charter schools. This terminology is used throughout the report to distinguish between the two kinds of public schools.

The battery of tests administered statewide in the spring of 2018 was officially named Student Assessment of Growth and Excellence (SAGE). The SAGE battery included a series of grade-level appropriate tests in three subject areas including:

1. English Language Arts (ELA)
2. Mathematics (Math)
3. Science

However, we decided not to include the Science tests in this analysis since they were not administered in some grades. This decision was made in consultation with representatives from the Office of the Legislative Auditor General. The consensus was that focusing on the ELA and

Math Tests would be sufficient to answer the research questions posed in this study. Hence, our analysis focused on the participation and nonparticipation rates on the ELA and Math tests.

The nonparticipation rate reported in this study generally includes students who were officially opted out from testing based on formal notification from a parent plus students who did not sit for one or both tests for some other reason such as those who legitimately missed school due to illness and those students who stayed out of school on their own initiative on days the tests were administered in order to avoid being tested. The files included about 400 (0.05 percent) students classified as "Student Refused to Test." We classified this category as a form of self opt out, but anecdotal reports from school administrators have led us to believe that the rate of self opt out likely exceeded this number. We suspect that many students who did not want to be tested found ways to avoid taking the tests without announcing their intention. Furthermore the variability within schools and between schools in how the various categories of untested students were coded leads us to question the accuracy and consistency of the way the categories were interpreted.

## METHOD AND RESULTS

### Research Questions 1 and 2

Before attempting to assess the impact of the missing data which resulted from nonparticipation on either the ELA or Math test, it is important to understand (a) how pervasive the missing data is, (b) how it is distributed across schools and across grade levels within schools, and (c) how it was correlated with other variables.

Some students responded to both the ELA and Math tests while others were examined on only one of the two tests, and others missed both tests. Table 1 reports the overall rate of participation and nonparticipation aggregated across grades 3–11. Each of the four rows in

**Table 1**

*Relative Frequency of the Four Different Patterns of Missing Scores on the ELA and Math Tests for All Students in Grades 3-11 by Type of School*

| Participation Pattern | Tests Taken | Students in District Schools[1] | | Students in Charter Schools | | Students in Virtual Schools[2] | | All Students Combined | |
|---|---|---|---|---|---|---|---|---|---|
| | | Number | Percent | Number | Percent | Number | Percent | Number | Percent |
| 1 | Both ELA and Math | 326,330 | 93.87 | 37,211 | 88.86 | 1,504 | 27.81 | 365,045 | 92.44 |
| 2 | ELA, but not Math | 3,418 | 0.98 | 332 | 0.79 | 42 | .78 | 3,792 | 0.96 |
| 3 | Math, but not ELA | 2,364 | 0.68 | 225 | 0.54 | 59 | 1.09 | 2,648 | 0.67 |
| 4 | Neither ELA or Math | 15,517 | 4.46 | 4,109 | 9.81 | 3,804 | 70.33 | 23,430 | 5.93 |
| Total | | 347,629 | 100.00 | 41,877 | 100.00 | 5,409 | 100.00 | 394,915 | 100.00 |

Notes:

[1]The term *district schools* refers to tradtional, neighborhood public schools which have a defined catchment area and are governed by the elected board of education in one of the 41 school districts in Utah.

[2]In reality, each virtual school is either an online charter school, or it is operated as an online school by a school district. However, the three categories of schools in this table are mutually exclusive meaning that students in virtual schools are counted separately and are not included in the counts for the charter or noncharter schools.

Table 1 represents a different pattern of participation. The first row (designated as Participation Pattern 1) represents students who sat for both the ELA test and the Math test, so this pattern is indicative of *complete participation*. As shown by the entry in the rightmost column of that row, 92.44 percent of the students in grades 3–11 statewide participated by taking both tests. In contrast, Participation Pattern 4 shows that 5.93 percent of the students statewide in grades 3-11 did not receive a score on either the ELA test or the Math test. This pattern is indicative of *complete nonparticipation*. The other two rows (Participation Patterns 2 and 3) in Table 1 represent different types of *partial nonparticipation*. Approximately 1 percent (0.96%) of the students statewide completed the ELA test but not the Math test, and two-thirds of one percent (0.67%) of the students completed the Math test but not the ELA test. Therefore, the overall

nonparticipation rate on the two tests was 7.56 percent. This total includes all students who did not participate in one or both tests.

However, these aggregated rates obscure the fact that the nonparticipation was higher among students in charter schools than in the district schools and even higher in virtual (online) schools. As reported in Table 1, the rate of nonparticipation on both tests (Pattern 4) was 4.46 percent in traditional district schools, 9.81 percent in charter schools, and 70.33 percent in virtual schools. This means that the complete nonparticipation is more than two times higher in charter schools than in district schools, and more than 15 times higher in virtual schools than in district schools. However, the combined rate of partial nonparticipation (Patterns 2 and 3 combined) is less than 2 percent and is essentially the same across all three types of schools.

Some of the online or virtual schools are sponsored by school districts, while others are online charter schools, but these three categories are considered to be mutually exclusive for the purposes of this study. Hence, the students in each of these three types of schools are counted separately in Table 1. However, when interpreting the nonparticipation rates reported in Table 1 readers should keep in mind that seven out of every eight students (88 percent) in 2018 were enrolled in district schools. Only about one-tenth of the students (10.6%) were enrolled in charter schools, and only 1.37 percent were enrolled in online schools. So, even though the nonparticipation rates are higher in charter schools and even higher in online schools, these two groups together accounted for only about one-eighth (12%) of the students in grades 3-11 statewide in the spring of 2018.

Nonparticipation rates for students enrolled in specific online schools are summarized in Table 2 for the English Language Arts Test and in Table 3 for the Mathematics Test. The letters

**Table 2**

*Percent of Nonparticipation in the 2018 Sage English Language Arts*
*by Online LEA and Grade-level Category*

| LEA Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Alpine District | 57.14 | 62.26 | NA |
| Athenian eAcademy | 48.96 | 52.38 | 48.28 |
| Canyons District | NA | NA | 43.18 |
| Davis District | 68.97 | NA | NA |
| Leadership Academy of Utal | 77.78 | 73.33 | 89.66 |
| Lumen Scholar Institute | 74.10 | 61.05 | 72.73 |
| Mountain Heights Academy | NA | 66.03 | 73.72 |
| Provo District | 93.18 | 95.92 | 91.67 |
| South Sanpete District | 0.00 | 0.00 | NA |
| Uintah District | 30.77 | 66.67 | NA |
| Utah Connections Academy | 52.41 | 49.72 | 51.89 |
| Utah Virtual Academy | 48.56 | 44.10 | 52.87 |
| Wasatch District | 100.00 | 100.00 | NA |
| Washington District | 95.86 | 95.09 | 90.32 |
| Weber District | 0.00 | NA | NA |

"NA" in each of these tables indicates that the test was not administered at the corresponding grade levels in that school. In most cases this means that online offerings are not available for students in that school for students in the grade levels marked "NA." Note that with only one exception, all of the "NA"s occur in conjunction with a school district. The only exception is in the column for grades 3-6 at Mountain Heights Academy on both the English Language Arts Test and on the Mathematics Test. This virtual school did not include students in grades 3-6 in the spring of 2018.

Inspection of the cell entries in Table 2 indicates that the percent of nonparticipation on the English Language Arts test ranges from zero to more than 90 percent and that most of the

reported percentages exceed 50 percent. The reported nonparticipation rates of 0.00 percent in grades 3-6 in the Weber School District and in grades 3-6 and 7-8 in the South Sanpete School District on the English Language Arts Test are noteworthy and commendable, but they appear to be outliers when compared to the nonparticipation rates in the other virtual schools. The only other nonparticipation rate reported in Table 2 that is less than 40 percent is the 30.77 percent rate reported for grades 3-6 in the Uintah School District. The three online institutions with the highest rates of nonparticipation are the Provo, Wasatch, and Washington School Districts. The online school in the Washington District serves students from many areas across the state. This may also be true of the online schools in some of the other districts. The online charter school with the largest percent on nonparticipation is the Leadership Academy of Utah.

Table 3 reports the nonparticipation rate on the Mathematics Test in online schools. On this test the percent of nonparticipation ranges from 0.00 percent in grades 3-6 in the Weber School District to 100 percent in grades 3-6 and 7-8 in the Wasatch School District and in grades 9-11 in the Provo District. Note that missingness exceeds 90 percent for all three grade levels in Provo and the nonparticipation rate is more than 88.5 percent in all three grade categories in the Washington School District.

Students enrolled in district schools and most charter schools generally have an opportunity to take the tests in the school which they attend, but students enrolled in virtual schools typically need to travel to a temporary testing center set up at various sites around the state. These testing centers may be located in a church building, in a conference room of a hotel, or in some other temporary location. Since there are usually only about six or eight of these sites in a given year, some online students may have to travel several hours or more to reach the testing center and may be faced with the need to stay overnight. Traveling to the testing location on the scheduled dates may be problematic for online students who reside in remote parts of the

**Table 3**

*Percent of Nonparticipation in the 2018 Sage Math Test by Online LEA and Grade Level Category*

| LEA Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Alpine District | 55.00 | 58.49 | NA |
| Athenian eAcademy | 46.84 | 52.38 | 29.17 |
| Canyons District | NA | NA | 54.88 |
| Davis District | 68.97 | NA | NA |
| Leadership Academy of Utah | 80.77 | 76.12 | 86.08 |
| Lumen Scholar Institute | 71.69 | 59.15 | 70.15 |
| Mountain Heights Academy | NA | 66.67 | 74.92 |
| Provo District | 93.19 | 95.58 | 100.00 |
| Uintah District | 23.08 | 36.36 | NA |
| Utah Connections Academy | 50.54 | 40.45 | 52.90 |
| Utah Virtual Academy | 49.35 | 44.82 | 54.15 |
| Wasatch District | 100.00 | 100.00 | NA |
| Washington District | 96.08 | 93.49 | 88.62 |
| Weber District | 0.00 | NA | NA |

state. This difficulty of accessing the tests is likely a factor that contributed to the high rate of nonparticipation among online students.

In addition to varying by type of school, the nonparticipation rate also varies across grade levels within schools and between geographic areas across the state as reported in the three tables included in the appendix of this report. Tables A1-A3 report the nonparticipation rates separately for school districts and charter schools and by three different grade-level categories.

Table A1 reports the nonparticipation rates for each of the 41 school districts in the state on both the ELA and Mathematics Tests. Nonparticipation was higher in five school districts

(Park City, Provo, Rich, Tintic, and Tooele) than in the other districts. The two highest nonparticipation rates are for grades 9-11 in the Garfield and Uintah Districts. Nonparticipation is relatively low for grades 3-6 and 7-8 in both of these districts, but much higher in grades 9-11 compared to other school districts. The high rates in grades 9-11 likely reflect a decision not to expect eleventh grade students to participate in the testing.

Tables A2 and A3 display how the nonparticipation rate varies across charter schools. Table A2 reports the percent on nonparticipation on the ELA Test, and Table A3 reports the nonparticipation rate on the Math Test. As previously reported in Table 1, nonparticipation tends to be higher in charter schools than in neighborhood schools. The data in Tables A2 and A3 support this earlier conclusion, but Tables A2 and A3 also reveal that there is much greater variability in nonparticipation from one charter school to another than among the school districts reported in Table A1.

In summary, the data reported in Tables 1-3 and in Tables A1-A3 (in the appendix) clearly show that the missing test scores are not distributed randomly across schools or across grades within schools. Nonparticipation occurs more frequently in online and charter schools than in district schools, and it also occurs more frequently in high school grades than in lower grade levels. Nonparticipation also varies greatly across geographic areas within the state. Although it occurs to some extent in almost all schools and in all areas of the state, there appears to be a contagion effect associated with nonparticipation in the sense that it tends to be concentrated in some schools and in some localities more than in others, and the incidence is higher among some subgroups than others.

As we stated previously, the type of missingness is of greater concern than the amount. To further explore the type of missingness in the data, we created a missingness indicator variable for the ELA Test (“ELAmiss”) and a separate missingness variable for the Math Test

("Mathmiss"). To create these indicators, each student who had a missing score on the ELA Test was assigned an ELAmiss code of "1," and each student who did not have a missing ELA score was coded as "0." Similarly, each student who had a missing score on the Math Test was assigned a Mathmiss code of "1," and each student who did not have a missing score on the Math Test was coded as "0." We then computed the correlation between each of these missingness indicators and various demographic variables that were in the data sets made available to us. These variables included (a) gender, (b), eligible or not eligible to receive free or reduced lunch, (c) Caucasian or nonCaucasian, and (d) student mobility (whether or not the student had been enrolled in the same school for 160 days or more during the current academic year).

Although the correlations are modest, the results showed that nonparticipation is significantly correlated with four demographic variables in addition to being related to the type of school in which a student was enrolled and the students' grade level in school.

1. gender (females were more likely to be nonparticipants)
2. socioeconomic status (students who did *not* qualify for free or reduced lunch were more likely to be nonparticipants)
3. ethnicity (Caucasian students were more likely to not participate in the testing than students who were not Caucasian)
4. mobility (students who were enrolled in their school for less than 160 days were more likely to be nonparticipants).

Therefore, nonparticipation is at least somewhat predictable using the type of school in which a student was enrolled, their grade level, and their score on each of the four variables in the above list.

**Research Question 3**

To address question 3 we limited the population studied to grades 3-8 since elementary schools and middle schools are most at risk for being classified as needing Comprehensive Support and Improvement (CSI). In addition, we excluded very small schools (i.e., those with less than 30 students enrolled) and schools that we believed would not be candidates for being identified for CSI such as (a) the Utah School for the Deaf and Blind and (b) Summit Academy High School which serves persons incarcerated at the Utah State Prison.

Modern methods for treating missing data make it possible to impute plausible estimates of missing scores and then substitute the resulting imputed values in place of the missing scores (Cole, 2008; Enders, 2010; Little & Rubin, 2020; Schafer, 1998; Schafer & Graham, 2002). We investigated Research Question 3 by constructing such a hypothetical data set containing imputed scores substituted in place of the missing test scores for the untested students. We then compared this counterfactual data with the real data set which included missing scores for the untested students.

For each student who had a missing score on either one or both of the ELA and Math tests we used multilevel, multiple imputation to obtain a plausible estimate of the score which they would have received had they completed the test. Multiple imputation was first proposed by Rubin (1978) and later more clearly articulated (Little & Rubin, 1987, 2002, 2020). Nowadays it is one of the two most widely used methods to predict missingness. The procedure is based on a regression framework. The predictor variables that we used included (a) the type of school in which the student was enrolled (district school, charter school, or online school), (b) grade level (c) gender (d) Caucasian or nonCaucasian, (e) socio-economic status (i.e., whether or not the student qualified to receive free or reduced lunch) and (f) the student's test score on the corresponding test from the previous year when available.

Multiple imputation procedures add a random value to each score predicted by the regression equation to account for the variability that would likely be present in the scores obtained by students who have the same level on each of the predictor variables. This additional term can be positive or negative. It is sampled from a normal distribution with a mean of zero and a variance equal to the variance of the residuals obtained from the regression model. Including this random element acknowledges that the imputed values all lie on the regression surface and consequently lack the variability or uncertainty that would be expected (Enders, 2010; Little & Rubin, 2020).

We used a modified multiple imputation procedure. The procedure still involved multiple imputations in the sense that we imputed ten different sets of plausible replacement values for each student who had a missing SAGE score. However, instead of computing parameter estimates (such as the mean and variance) ten different times (once for each set of set of imputed scores) and then computing an average of the resulting parameters, we used a more direct procedure. For each student who had a missing score, we averaged the ten imputed scores across the ten sets of repeated estimates, to obtain a plausible estimate of the missing score. We used this modified, but more direct procedure, for two reasons:

1. Since the dataset contained the entire population of interest, no inferential tests were required, and
2. Since the prescribed cutoff criterion for classifying schools into CSI status is focused on the mean of each school, the variance in student scores is not of major interest.

Furthermore, the variation of the summary statistics was minimal (< .01 of one standard deviation) indicating that the results are robust to the choice of aggregation method.

We also conducted a logistic regression analysis that included the same covariates that were used in the imputation procedure in order to estimate the strength of the associations predicting

missingness. The results showed a small, but statistically significant, relationship (Cox & Snell R square =.128).

For both the ELA test and the Math test, the resulting augmented data included the observed score for each student who was originally tested and the mean imputed score for each student who was not tested. These hypothetical data sets provided the counterfactual data that we analyzed to address research question 3.

We used the rules specified in the "Utah Accountability Technical Manual" (https://www.schools.utah.gov/file/70235d75-cf35-4e04-9d2b-34ff388968b5) to determine the identification status of each elementary and middle school. On page 60, this manual states "After calculations are completed, a ranked list of all schools will be produced, with ranking based on the percent of total points earned out of total points possible for all available indicators for a school." On page 33 of the same manual, it states that the lowest performing five percent of Title I schools in the state are to be identified for Comprehensive Support and Improvement (CSI). In other words, the lowest performing five percent in the ranked list of schools are classified as needing CSI.

Since the data set which we analyzed did not include all of the indicator variables that are used to identify CSI schools, we focused our analysis solely on the total number of points that a school could earn from the performance of its students on the SAGE ELA and Math tests. For each of 740 elementary and middle schools in the population studied, we followed the procedures described in the technical manual. We added all the scores earned by the students in that school on both the SAGE Math test and the SAGE ELA test. Then for each school, we divided the *total points earned* by the *total points possible* (i.e., the total points that could have been earned from the ELA and Math tests) to obtain that school's percentage of the total possible points. We followed this procedure twice in order to obtain two scores for each school: first, a

percent score based on the original file of reported data that did not include test scores for untested students, and second, a percent score based on data from the augmented file that included imputed scores for the untested students in addition to the scores obtained by the tested students. Two lists of rank ordered schools were then produced: one ranked list based on the original reported data, and a second list based on the file in which missing scores were replaced with imputed scores.

We used two different methods to compare the two ranked lists. One method was to compute the correlation between the two lists. The resulting correlation coefficient was .977. This high correlation indicates that the relative standing of individual schools in the two lists is very similar with few exceptions. In other words, schools that are ranked high in one list tend to have a similar high ranking the other list, and schools that have a low rank in one list tend to have a similar low rank in the other list. Similarly, schools that are about average in one list are about average in the other list.

When schools were ranked based on the data without imputed scores for the untested students, the cut-off for identifying the lowest performing five percent of all the schools was 51.20 percent of the total possible points. In contrast, the cut-off for identifying the lowest five percent of all schools based on the data that included imputed scores for untested students was 51.27%.

Table 4 summarizes the extent to which schools were classified consistently based on the two different data sets. The two interior rows in Table 4 display how the schools were classified based on the original, reported data which included missing scores for untested students. On the other hand, the two columns in Table 4 show how the schools were classified based on the data set that included imputed scores for the untested students. In accord with the rules, under both

**Table 4**

*Number of Schools Classifed as Performing Adequately and Those Classified as Needing Improvement Based on the Actual Data Versus the Number Classified in Each of Those Categories Based on the Counterfactual Data*

| | | School Status Based on Counterfactual Data | | |
|---|---|---|---|---|
| | | Adequately Performing Schools | Schools Targeted for Improvement | Total Number of Schools |
| School Status Based on Reported Data | Adequately Performing Schools | 701 (94.73)% | 2 (0.27%) | 703 (95.0%) |
| | Schools Targeted for Improvement | 2 (0.27%) | 35 (4.73%) | 37 (5.0%) |
| | Total Number of Schools | 703 (95.0%) | 37 (5.0%) | 740 (100.0%) |

analytic scenarios, 703 (95%) of the schools were classified in the satisfactory category, and 37(5%) were identified for CSI status. However, the schools classified in each category under the two scenarios were not exactly the same.

A total of 701 (94.73%) schools were classified in the adequately performing category under both analytic scenarios as indicated in the upper left shaded cell, and 35 (4.73%) were classified in the underperforming category under both scenarios. Hence, the two shaded cells in the main diagonal of Table 4 report the number of schools that were consistently classified across both data sets. Adding the numbers in these two cells together indicates that 736

(99.46%) of the 740 schools were consistently classified. The two off-diagonal cells in the body of the table indicate the number of schools that were inconsistently classified. Only 4 (0.54%) of the 740 schools were inconsistently classified by the two different data sets. Two schools that were originally classified in the acceptable category were subsequently classified in the CSI category when the imputed scores were included in the analysis. Conversely, two other schools that were originally identified as CSI schools were re-classified in the adequately performing category when the imputed scores were included.

From a statistical point of view, the effect of missing data on the CSI status of schools based on the 2018 SAGE tests was minimal. Replacing the missing data with our best estimates of the scores the nonparticipating students likely would have received if they had been tested did not change the CSI status of 99.46% of the schools. However, from a practical point of view, the evidence indicates that the missing scores did make a difference for four schools, but those four represent only about one-half of one percent of all the elementary and middle schools in the state.

**Research Question 4**

To investigate the anecdotal claim that higher achieving students have a higher rate of nonparticipation than other students we merged the tenth-grade 2018 and seventh-grade 2015 statewide results for SAGE Math and SAGE ELA test results. In this way students were identified in the same file who had taken both the seventh- and tenth-grade tests. We selected tenth-graders because they have the most consistent and robust high school sample. In the eleventh grade fewer students were tested and there is increased variability in which tests are taken by which students. We selected seventh-grade tests because seventh grade is the first year of secondary testing and would be the most similar classroom and testing context to the tenth-

grade testing context. Furthermore, it was likely that students who opted out in tenth grade were likely to have opted out in ninth grade. Using the seventh-grade tests would enhance the probability of finding students that eventually opted out of tenth grade testing but who had taken the tests in seventh grade. Using the merged files, we then identified students who opted out of the tenth-grade test in 2018, tagged them as a separate group, and compared their performance on the seventh-grade test in 2015 to the performance of all other seventh graders in 2015.

The results are summarized in Table 5. These results indicate that the tenth graders who *opted out of* the 2018 SAGE tests had significantly lower means on both 2015 seventh-grade tests than the students who *did participate* on the 2018 SAGE tests. This finding runs counter to the anecdotal claim that higher achieving students tend to opt out at a higher rate than lower achieving students.

**Table 5**

*Mean Scaled Scores on the 2015 Seventh-grade SAGE Tests by Students Who Later Became Participants or Nonparticipants on the 2018 SAGE Tests When They Were Tenth Graders*

| | 2018 SAGE Participants | | | 2018 SAGE Nonparticipants | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Test | Number of Students | Seventh grade Mean | Standard Deviation | Number of Students | Seventh grade Mean | Standard Deviation | Difference in Means | *t* | Probability | Effect Size |
| ELA | 42,177 | 433.215 | 77.085 | 2,197 | 423.334 | 76.967 | 9.881 | 5.858 | 0.001 | .1282 |
| Math | 41,564 | 440.326 | 58.114 | 2,021 | 431.693 | 50.963 | 8.633 | 7.383 | 0.001 | .1682 |

On both the ELA test and the Math test, the difference between the mean scores of students who participated in the 2018 tests and those who did not participate were statistically significant. The tenth graders who *did not participate* in the 2018 SAGE test had significantly lower means on both 2015 seventh-grade tests than the tenth-graders who *did participate* on the 2018 SAGE tests.

The magnitude of the observed differences reported in Table 5 is described by the effect sizes reported in the far-right column of the table. On the average, the students who did not participate in the 2018 ELA test scored one eighth of a standard deviation lower than all other seventh graders on the 2015 ELA test ($d$ = .1282). Similarly, on the average, the students who did not participate in the 2018 Math test scored one-sixth of a standard deviation lower than all other seventh graders on the 2015 Math test ($d$ = .1682).

These findings run counter to the anecdotal claim that higher achieving students opt out at a higher rate than lower achieving students. In fact, the opposite was true for tenth graders. Students who were relatively less proficient in seventh grade had a greater tendency to not participate in the SAGE tests when they became sophomores. We do not know to what extent the differences observed between the tested and untested tenth graders is characteristic of students in other grades.

We also compared the tenth-grade participants and nonparticipants in terms of demographic variables. The difference between the groups was not statistically significant in terms of the family income variable in the files we had, but the two groups did differ significantly in terms of gender and minority status on both the ELA and Math tests. The proportion of tenth-grade females who did not participate was significantly greater than the proportion of their male counterparts who did not participate. Similarly, the proportion of tenth-grade majority students who did not participate was significantly greater than the proportion of tenth-grade minority students who did not participate.

**Research Question 5**

Given the limitations of the data available to us, it was not possible to obtain an adequate answer to the question about how the low-stakes nature of the SAGE tests may have influenced students' degree of motivation and engagement when responding to the tests. The data consisted

of scores from four local education agencies each of which had administered another achievement test in addition to the SAGE tests. However, we were able to address this question indirectly and obliquely by analyzing scores obtained from two Utah school districts and two charter schools each of which had administered a different test of their own choosing to some of their students in addition to the SAGE tests. Is this report, these four LEAs are referred to Districts A and B and Charter Schools A and B.

District A administered the *Galileo* test under high stakes conditions, and District B administered the *Aleks Diagnostic* test under high stakes conditions. In addition, we were given data from one charter school which had administered the MAP test marketed by the Northwest Evaluation Association, and data on another unknown test used by a different charter school. For each of these data sets we computed the correlation between students' scores on the SAGE tests and their scores on the district-administered, high-stakes test. The constructs assessed by these other tests are somewhat similar to the student achievement constructs which the SAGE tests are intended to assess. Nevertheless, each of these additional assessments is a unique test with its own distinctive items and content coverage. Since disengagement is more likely to appear among high school students than among students in lower grades, we analyzed only data for students enrolled at the high school level. The correlations between scores on these different tests and the two SAGE tests ranged between .63 and .82.

The resulting correlation coefficients provide some evidence that students tend to have about the same relative standing on the SAGE tests administered under low stakes conditions as they do on these alternative tests administered under conditions where the perceived stakes are higher. However, in our judgment the observed correlations do not directly address the question as to whether students lack sufficient motivation to perform well on the SAGE Tests. The results of these analyses should be considered tentative at best. Further investigation is needed.

When teachers, administrators, policy makers, and other stakeholders interpret test scores from classes, schools, or states as a basis for making judgments about the effectiveness of teachers or schools, they generally assume that the students who took the tests invested effort in trying to solve the problems and answer the questions on the test correctly and that their responses are indicative of what the students actually know and can do (Wolf & Smith, 1995). But numerous researchers have shown that students' scores on a test are a product of their test-taking motivation as well as their knowledge and skill (Eklof, 2010; Wise, 2015). This situation presents a problem. In the words of Howard Wainer. "Often what we are trying to measure is examinees' proficiency when they are really trying, [But] If a test doesn't count for specific individuals, how can we be sure that they are trying as hard as they might if it mattered" (1993, p. 12).

Numerous scholars have demonstrated that test-taking motivation matters and that students respond differently when they perceive that a test does not have consequences for them personally (Finn, 2015; Knekta, 2017; Penk & Richter, 2017). This means that decisions about the effectiveness of teachers and schools based on students' scores on low-stakes tests are likely to be less valid (Finney, Sundre, Swain, & Williams 2016; Wise, 2015; Wise & DeMars, 2005)

Steven Wise and his associates (Wise, 2015; Wise & DeMars, 2005; Wise & Kingsbury, 2016; Wise & Kong, 2005) have completed a series of studies in which they investigated various ways of collecting evidence of examinees' lack of motivation in responding to low-stakes tests. They tried using self-reported measures of effort and engagement collected by administering a questionnaire to students immediately after completing a test and found that this method was subject to intentional distortion by the students and did not yield a valid measure of students' motivation or engagement. They also reported evidence that a student's effort in responding to a test is not a static characteristic. Instead, test-taking motivation is a dynamic characteristic which

varies from one test item to another during the time that the test is being administered. Furthermore, self-reported measures are not sufficiently sensitive to these changes in level of motivation as students encounter different items. Some scholars have advocated the use of person-fit statistics generated by item response theory models as an alternative way of identifying students who lack motivation, but Wise and his collaborators pointed out that these statistics are influenced by a number of aberrant behaviors other than lack of motivation.

Because they were dissatisfied with the existing ways of assessing the amount of effort students devote to responding to test items, Wise and his associates hypothesized that students who lack motivation to perform well on a test item engage in rapid-guessing behavior. Hence, instead of spending the time needed to carefully read the stem of the question and the various options, unmotivated students quickly respond by guessing. Computer-administered tests are especially vulnerable to random guessing since they do not allow examinees to skip items. Consequently, Wise and his colleagues developed an index of student effort based on the amount of time measured in seconds that students spend on each item in a test. In reality, this index is a proxy measure of test-taking effort, but Wise and his collaborators have provided evidence that it is a more valid indicator of students' motivation and engagement in responding to the tests than person fit statistics or responses to a self-report questionnaire.

Other researchers have also proposed ways to identify examinees who engage in noneffortful behavior when responding to low-stakes tests. For instance, Guo, Rios, Haberman, Liu, Wang, & Paek (2016) proposed an improved way of implementing the response time index proposed by Wise and his colleagues. Other researchers have utilized alternative methodologies to identify and describe differences in students' degree of test-taking motivation including latent growth modeling (Barry & Finney, 2016) and mixture modeling (Barry, Horst, Finney, Brown, & Kopp, 2010). In order to employ these methods, researchers need access to item response data

from the test and/or a measure of the amount of time students expend in responding to the various items. If the necessary data can be made available, there is no lack of available methodological tools that can be used to identify examinees who respond in a noneffortful manner. It may be helpful to utilize more than one of these methods in the same study.

Since both the item response data and the elapsed-time-per-item information are routinely collected when tests are administered online, obtaining this data from state administered tests in a deidentified form should not be difficult. If the state desires to pursue such a study, detailed arrangements for obtaining the data should be made in advance of the testing occasion.

At present, we do not know to what extent the lack of test-taking motivation poses a threat to the validity of inferences about school effectiveness that are likely to be made by policy makers and other stakeholders based on scores obtained from state-administered tests in Utah. We have evidence from previous research completed in other contexts that when students are asked to take tests that have no direct consequences for them personally, instead of exerting the effort necessary to determine the best answer to each question they often resort to rapid-guessing, especially when they encounter difficult problems or questions. Although we suspect that such noneffortful behavior is more prevalent among students in secondary schools than in elementary schools, we really do not know how prevalent it is. If it is widespread, then this lack of motivation is likely to be a greater validity threat than nonparticipation due to opt out. However, without further research based on Utah students' responses to the Utah tests we have no way of knowing how much of a validity threat is posed by this potential problem. We believe that this is a critical issue which warrants further investigation.

## LIMITATIONS

The analyses and conclusions reported here are based on the assumption that the multiple imputation procedure we used produced a plausible estimate of the missing scores. We are confident that the procedures we used produced the best estimates possible given the available predictors, but readers should keep in mind that the imputed scores are estimates. They are not wild guesses, but they are estimates. The estimates we obtained were predicted from several available variables using a regression procedure that was replicated multiple times. This imputation procedure would likely have produced more precise estimates if stronger predictor variables had been available.

## CONCLUSIONS

The following conclusions are offered based on the results of our analyses.

1. Missing scores on the 2018 SAGE tests were not a random occurrence. Opting out occurred more frequently in charter schools and in online schools than in district schools, and it occurred relatively more frequently in older grade levels than in younger grade levels. In addition, it is related to several demographic variables, although the correlations are weak, but statistically significant. Therefore, nonparticipation is at least somewhat predictable.
2. Overall, opting out had a minimal effect in terms of decisions about which schools would be identified for Comprehensive Support and Improvement. Because a school has to be below the cut-off for three consecutive years before being classified in this category, the risk of false- positive decisions is extremely small.
3. We did not find evidence to support the anecdotal claim that higher-achieving students opt out more frequently than their lower-achieving counterparts. These results are

contrary to the anecdotal claims that higher achieving students have a greater tendency to opt out than lower achieving students.

4. Based on this current analysis, opt out has a minimal threat to the validity of inference based on the observed test scores. A greater threat may be due to the lack of motivation on the part of students since the tests do not have personal consequences for them. If this lack of motivation due to the perceived low-stakes nature of the tests is widespread, then the Utah Legislature may have inadvertently hurt their own cause by implementing the law [53E-4-303(4)(a)] which limits the use of the test scores in ways that would have consequences for students.

## RECOMMENDATIONS

We recommend that the following actions be taken based on the results of our investigation and analyses.

1. Each school and local education agency in the state, especially charter schools and online schools, should be given explicit directives and guidelines regarding their responsibilities for administering the state tests and processing opt-out requests from parents.
2. Uniform rules for classifying and coding absences from the tests should be implemented. Local school administrators, especially those who administer the tests, should be trained in implementing these rules.
3. Further research should be conducted to investigate the degree to which students fail to take the tests seriously and give their best efforts in responding. This research should be based on analysis of students' response time per item and the accuracy of

their responses. It should not only attempt to document the degree of disengagement but also examine the factors that contribute to it.

4. Conduct a follow-up study in future years with the RISE tests to determine to what extent the findings of this initial study are generalizable.

## REFERENCES


Allison, P.D. (2002). *Missing data*. Thousand Oaks, CA: Sage Publications.

Barry, C.L. & Finney, S.J. (2016). Modeling change in effort across a low-stakes testing session: A latent growth curve modeling approach. *Applied Measurement in Education, 29,* 46-64.

Barry, C.L., Horst, S.J., Finney, S.J., Brown, A.R., & Kopp, J.P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International Journal of Testing, 10*, 342-363.

Cole, J.C. (2008). How to deal with missing data: Conceptual overview and details for implementing two modern methods. In J.W. Osborne (Ed.), *Best practices in quantitative* methods (pp. 214-238). Thousand Oaks, CA.

Eklof, H. (2010). Skill and will: Test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice, 17*, 345-356.

Enders, C.K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Finney, S.J., Sundre, D.L., Swain, M.S., & Williams, L.M. (2016). The validity of value-added estimates from low-stakes testing contexts: The impact of change in test-taking motivation and test consequences. *Educational Assessment, 21*, 60-87.

Guo, H., Rios, J.A., Haberman, S., Liu, O.L., Wang, J., & Paek, I. (2016). A new procedure for detection of students' rapid guessing responses using response time. *Applied Measurement in Education, 29*, 173-183.

Knekta, E. (2017). Are all pupils equally motivated to do their best on all tests" Differences in reported test-taking motivation within and between tests with different stakes. *Scandinavian Journal of Educational Research, 61*, 95-111.

Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York, NY: John Wiley.

Little, R.J.A. & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley.

Little, R.J.A. & Rubin, D.B. (2020). *Statistical analysis with missing data* (3rd ed.). Hoboken, NJ: John Wiley.

Penk C. & Richter, D. (2017). Change in test-taking motivation and its relationship to test performance in low-stakes assessments. *Educational Assessment, Evaluation, and Accountability, 29*, 55-79

Rubin, D.B. (1976). Inference and missing data. *Biometrika, 63*, 581-592.

Rubin, D.B. (1978). Multiple imputations in sample surveys. In *Proceedings of the Survey Research Methods Section, 1978*, 20-34. American Statistical Association.

Schafer, J.L. (1998). *Multiple imputation: A primer*. *Statistical Methods in Medical Research, 8*, 3-15.

Schafer, J. & Graham, J. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement, 30*, 1-21.

Wise, S.L. (2015). Effort analysis: Individual score validation of achievement test data. *Applied Measurement in Education, 28*, 237-252.

Wise, S.L. & DeMars, C.E. (2005). Examinee motivation in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 38*, 1-17.

Wise, S.L. & DeMars, C.E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment, 15*, 27-41.

Wise, S.L. & Kingsbury, G.G. (2016). Modeling student test-taking motivation in the context of an adaptive achievement test. *Journal of Educational Measurement, 53*(1), 86-105.

Wise, S.L. & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183.

Wolf, L.F. & Smith, J.K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*, 227-242.

## APPENDIX

**Table A1**

*Percent of Nonparticipation on the ELA and Math Tests by School District and Grade Level Category*

| | English Language Arts Test | | | Mathematics Test | | |
|---|---|---|---|---|---|---|
| School District | Grades 3-6 | Grades 7-8 | Grades 9-11 | Grades 3-6 | Grades 7-8 | Grades 9-11 |
| Alpine | 4.75 | 6.13 | 7.56 | 4.60 | 5.86 | 9.38 |
| Beaver | 4.02 | 3.54 | 2.67 | 3.60 | 3.16 | 2.68 |
| Box Elder | 1.63 | 3.96 | 9.58 | 1.86 | 3.01 | 4.61 |
| Cache | 2.67 | 3.68 | 3.67 | 2.98 | 4.02 | 4.85 |
| Canyons | 4.10 | 8.21 | 8.43 | 3.81 | 8.75 | 8.77 |
| Carbon | 9.83 | 4.50 | 6.04 | 2.77 | 4.70 | 5.63 |
| Daggett | 0.00 | 0.00 | 4.88 | 0.00 | 3.23 | 9.76 |
| Davis | 3.89 | 5.78 | 6.63 | 3.72 | 6.08 | 8.01 |
| Duchesne | 3.32 | 4.63 | 3.51 | 3.49 | 4.81 | 3.62 |
| Emery | 2.31 | 3.62 | 2.29 | 1.31 | 1.98 | 2.01 |
| Garfield | 4.33 | 0.67 | 28.65 | 2.31 | 1.34 | 2.76 |
| Grand | 0.47 | 1.16 | 6.33 | 0.71 | 1.92 | 6.54 |
| Granite | 2.51 | 3.97 | 6.88 | 2.61 | 3.76 | 7.74 |
| Iron | 2.53 | 4.03 | 6.92 | 2.67 | 4.33 | 8.23 |
| Jordan | 4.44 | 5.67 | 10.69 | 4.84 | 6.05 | 10.41 |
| Juab | 5.33 | 3.90 | 9.50 | 5.31 | 5.73 | 8.16 |
| Kane | 2.53 | 2.46 | 3.91 | 3.18 | 2.46 | 3.91 |
| Logan City | 2.72 | 9.49 | 17.04 | 2.81 | 8.88 | 16.92 |
| Millard | 1.75 | 5.71 | 4.47 | 1.86 | 5.92 | 4.95 |
| Morgan | 3.74 | 5.94 | 4.45 | 3.43 | 5.93 | 5.36 |
| Murray | 2.39 | 3.57 | 2.92 | 1.86 | 4.49 | 3.65 |
| Nebo | 3.47 | 6.28 | 7.24 | 3.86 | 5.87 | 6.73 |
| North Sanpete | 1.51 | 5.01 | 2.85 | 1.64 | 3.20 | 4.27 |
| North Summit | 2.69 | 3.33 | 2.94 | 2.69 | 4.00 | 3.70 |
| Ogden City | 2.06 | 4.32 | 6.17 | 2.85 | 5.00 | 9.37 |
| Park City | 9.86 | 20.60 | 23.17 | 9.52 | 20.11 | 24.71 |
| Piute | 1.59 | 0.00 | 7.14 | 1.59 | 0.00 | 5.36 |
| Provo City | 20.50 | 27.36 | 12.96 | 3.83 | 21.10 | 11.85 |
| Rich | 19.11 | 23.29 | 18.84 | 19.11 | 21.92 | 18.84 |
| Salt Lake City | 2.11 | 3.90 | 7.61 | 2.25 | 4.04 | 7.07 |
| San Juan | 2.10 | 3.25 | 4.70 | 1.98 | 2.81 | 3.70 |
| Sevier | 1.33 | 4.08 | 5.37 | 1.17 | 4.09 | 4.04 |
| South Sanpete | 5.19 | 4.84 | 7.97 | 5.21 | 4.44 | 11.75 |
| South Summit | 5.23 | 5.88 | 8.37 | 4.29 | 5.88 | 8.16 |
| Tintic | 15.94 | 26.67 | 24.44 | 15.94 | 24.44 | 25.53 |
| Tooele | 16.93 | 16.54 | 21.45 | 18.99 | 13.08 | 12.10 |
| Uintah | 3.59 | 4.92 | 30.07 | 3.98 | 4.55 | 5.07 |
| Wasatch | 2.64 | 5.52 | 4.38 | 2.59 | 5.99 | 5.01 |
| Washington | 9.16 | 11.88 | 9.74 | 5.39 | 6.07 | 7.86 |
| Wayne | 4.20 | 11.76 | 14.29 | 6.72 | 11.76 | 17.39 |
| Weber | 3.04 | 5.98 | 8.03 | 2.84 | 5.63 | 6.92 |

**Table A2**

*Percent of Nonparticipation on the ELA Test by Charter School and Grade Level*

| School Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Academy for Math Engineering & Science | NA | NA | 3.31 |
| American Academy of Innovation | 43.14 | 39.45 | 64.71 |
| American International School of Utah | 22.36 | 20.89 | 15.61 |
| American Leadership Academy | 3.70 | 5.96 | 6.69 |
| American Preparatory Academy | 4.11 | 5.44 | 7.52 |
| Ascent Academies of Utah | 4.93 | 4.19 | 9.48 |
| Athlos Academy of Utah | 5.10 | 5.00 | NA |
| Bear River Charter School | 8.57 | 8.11 | NA |
| Beehive Science & Technology Academy | 3.92 | 1.72 | 4.17 |
| Bonneville Academy | 6.07 | 10.81 | NA |
| C.S. Lewis Academy | 3.52 | NA | NA |
| Canyon Grove Academy | 43.37 | 58.82 | NA |
| Canyon Rim Academy | 9.33 | NA | NA |
| Channing Hall | 4.42 | 4.90 | NA |
| City Academy | NA | 17.31 | 12.96 |
| Davinci Academy | 35.18 | 10.80 | 11.86 |
| Dixie Montessori Academy | 6.77 | 9.09 | NA |
| Dual Immersion Academy | 0.50 | 1.30 | NA |
| Early Light Academy at Daybreak | 4.41 | 6.99 | 5.15 |
| East Hollywood High | NA | NA | 11.11 |
| Edith Bowen Laboratory School | 1.97 | NA | NA |
| Endeavor Hall | 0.97 | 2.33 | NA |
| Entheos Academy | 2.61 | 3.02 | 2.63 |
| Esperanza School | 0.82 | NA | NA |
| Excelsior Academy | 6.64 | 10.53 | NA |
| Fast Forward High | NA | NA | 16.35 |
| Franklin Discovery Academy | 14.56 | NA | NA |
| Freedom Preparatory Academy | 8.26 | 10.64 | 6.61 |
| Gateway Preparatory Academy | 35.06 | 38.13 | NA |
| George Washington Academy | 9.69 | 13.87 | NA |
| Good Foundations Academy | 5.60 | NA | NA |
| Greenwood Charter School | 4.79 | 20.00 | NA |
| Guadalupe School | 1.37 | NA | NA |
| Hawthorn Academy | 3.30 | 2.79 | 0.00 |
| Highmark Charter School | 0.44 | 2.44 | 0.00 |

(Table continues)

**Table A2 continued**

| School Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Intech Collegiate High School | NA | NA | 4.13 |
| Itineris Early College High | NA | NA | 5.60 |
| Jefferson Academy | 7.89 | NA | NA |
| John Hancock Charter School | 9.26 | 50.00 | NA |
| Karl G. Maeser Preparatory Academy | NA | 57.21 | 64.25 |
| Lakeview Academy | 12.34 | 12.63 | 6.02 |
| Leadership Learning Academy | 1.85 | NA | NA |
| Legacy Preparatory Academy | 5.45 | 8.64 | 1.30 |
| Lincoln Academy | 10.37 | 5.48 | 6.25 |
| Mana Academy Charter School | 3.25 | 0.00 | 0.00 |
| Maria Montessori Academy | 10.43 | 22.67 | 24.00 |
| Merit College Preparatory Academy | NA | 19.00 | 10.66 |
| Moab Charter School | 1.54 | NA | NA |
| Monticello Academy | 2.72 | 4.76 | 0.00 |
| Mountain West Montessori Academy | 13.59 | 13.25 | 13.33 |
| Mountainville Academy | 9.76 | 20.17 | 14.29 |
| Navigator Pointe Academy | 8.17 | 13.10 | 25.00 |
| No. UT. Acad. for Math Engineering & Science | NA | NA | 46.75 |
| Noah Webster Academy | 4.18 | NA | NA |
| North Davis Preparatory Academy | 6.10 | 10.86 | 5.19 |
| North Star Academy | 3.37 | 10.08 | 7.14 |
| Odyssey Charter School | 9.26 | NA | NA |
| Ogden Preparatory Academy | 2.40 | 2.54 | 0.00 |
| Open Classroom | 10.97 | 14.29 | NA |
| Pacific Heritage Academy | 5.29 | 7.35 | NA |
| Paradigm High School | NA | 31.40 | 37.50 |
| Pinnacle Canyon Academy | 17.88 | 13.25 | 20.90 |
| Pioneer High School for the Performing Arts | NA | NA | 8.00 |
| Promontory School of Expeditionary Learning | 5.88 | 4.49 | NA |
| Providence Hall | 6.48 | 13.56 | 3.56 |
| Quest Academy | 3.21 | 7.02 | 17.07 |
| Ranches Academy | 5.00 | NA | NA |
| Reagan Academy | 2.29 | 4.55 | NA |
| Real Salt Lake Academy High School | NA | NA | 16.44 |
| Renaissance Academy | NA | NA | NA |
| Rockwell Charter High School | NA | 16.05 | 23.21 |
| Roots Charter High School | NA | NA | 67.21 |
| Salt Lake Arts Academy | 4.55 | 3.17 | NA |

(Table continues)

**Table A2 continued**

| School Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Salt Lake Center for Science Education | 4.29 | 1.80 | 7.89 |
| Salt Lake School for the Performing Arts | NA | NA | 14.77 |
| Scholar Academy | 8.54 | NA | NA |
| Soldier Hollow Charter School | 8.44 | 0.00 | NA |
| Spectrum Academy | 33.33 | 37.93 | 30.47 |
| St. George Academy | NA | 18.84 | 14.10 |
| Success Academy | NA | NA | 10.05 |
| Summit Academy | 4.90 | 8.15 | NA |
| Syracuse Arts Academy | 4.86 | 5.41 | 6.67 |
| Terra Academy | 8.21 | 16.33 | 6.80 |
| The Center for Creativity, Innovation and Discovery | 18.82 | NA | NA |
| Thomas Edison | 8.30 | 9.96 | 4.00 |
| Timpanogos Academy | 73.43 | NA | NA |
| Treeside Charter School | 8.97 | NA | NA |
| Tuacahn High School for the Performing Arts | NA | NA | 12.43 |
| Uintah River High | NA | NA | 51.52 |
| Utah Career Path High School | NA | NA | 24.32 |
| Utah County Academy of Science | NA | NA | 4.48 |
| Utah International Charter School | NA | 10.00 | 10.23 |
| Utah Military Academy | NA | 11.86 | 17.94 |
| Valley Academy | 13.75 | 12.00 | NA |
| Vanguard Academy | NA | 0.00 | 1.08 |
| Venture Academy | 4.35 | 9.73 | 10.59 |
| Vista at Entrada School of Performing Arts and Technology | 9.93 | 8.63 | NA |
| Voyage Academy | 3.91 | NA | NA |
| Walden School of Liberal Arts | 11.48 | 21.82 | 42.86 |
| Wallace Stegner Academy | 2.80 | 1.18 | NA |
| Wasatch Peak Academy | 5.81 | NA | NA |
| Wasatch Waldorf Charter School | 17.18 | 20.75 | NA |
| Weilenmann School of Discovery | 5.57 | 17.36 | NA |
| Winter Sports School | NA | NA | 5.17 |

Table A3

*Percent of Nonparticipation on the Math Test by Charter School and Grade Level*

| School Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Academy for Math Engineering & Science | NA | NA | 4.64 |
| American Academy of Innovation | 36.73 | 37.38 | 65.56 |
| American International School of Utah | 20.81 | 19.20 | 16.19 |
| American Leadership Academy | 3.71 | 4.38 | 6.13 |
| American Preparatory Academy | 4.11 | 9.02 | 10.36 |
| Ascent Academies of Utah | 5.17 | 3.91 | 10.62 |
| Athlos Academy of Utah | 5.31 | 5.00 | NA |
| Bear River Charter School | 7.14 | 8.11 | NA |
| Beehive Science & Technology Academy | 3.92 | 2.56 | 5.48 |
| Bonneville Academy | 5.73 | 10 | NA |
| C.S. Lewis Academy | 3.50 | NA | NA |
| Canyon Grove Academy | 43.37 | 60.00 | NA |
| Canyon Rim Academy | 9.33 | NA | NA |
| Channing Hall | 4.78 | 4.85 | NA |
| City Academy | NA | 15.69 | 4.92 |
| Davinci Academy | 34.35 | 10.34 | 13.14 |
| Dixie Montessori Academy | 7.25 | 9.09 | NA |
| Dual Immersion Academy | 0.99 | 2.60 | 6.19 |
| Early Light Academy at Daybreak | 4.67 | 8.11 | 20.00 |
| East Hollywood High | NA | NA | NA |
| Edith Bowen Laboratory School | 1.97 | NA | NA |
| Endeavor Hall | 1.46 | 2.33 | 2.63 |
| Entheos Academy | 2.61 | 3.50 | NA |
| Esperanza School | 0.82 | NA | NA |
| Excelsior Academy | 6.32 | 11.05 | 26.00 |
| Fast Forward High | NA | NA | NA |
| Franklin Discovery Academy | 14.15 | NA | 5.83 |
| Freedom Preparatory Academy | 8.44 | 10.64 | NA |
| Gateway Preparatory Academy | 34.85 | 38.85 | NA |
| George Washington Academy | 9.47 | 13.87 | NA |
| Good Foundations Academy | 5.60 | NA | NA |
| Greenwood Charter School | 4.81 | 20.00 | NA |
| Guadalupe School | 2.05 | NA | 0.00 |
| Hawthorn Academy | 3.63 | 2.86 | 0.00 |
| Highmark Charter School | 0.44 | 1.95 | 2.08 |

(Table continues)

**Table A3 continued**

| School Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Intech Collegiate High School | NA | NA | 6.40 |
| Itineris Early College High | NA | NA | NA |
| Jefferson Academy | 7.89 | NA | NA |
| John Hancock Charter School | 8.33 | 50 | 62.02 |
| Karl G. Maeser Preparatory Academy | NA | 55.78 | 9.30 |
| Lakeview Academy | 12.09 | 14.14 | NA |
| Leadership Learning Academy | 1.86 | NA | 2.56 |
| Legacy Preparatory Academy | 5.24 | 9.05 | 9.47 |
| Lincoln Academy | 10.64 | 7.34 | 0.00 |
| Mana Academy Charter School | 4.07 | 0.00 | 24.00 |
| Maria Montessori Academy | 10.43 | 24.00 | 7.50 |
| Merit College Preparatory Academy | NA | 18.27 | NA |
| Moab Charter School | 7.69 | NA | 1.64 |
| Monticello Academy | 3.40 | 5.44 | 15.38 |
| Mountain West Montessori Academy | 14.08 | 12.05 | 14.29 |
| Mountainville Academy | 9.97 | 20.34 | 25.00 |
| Navigator Pointe Academy | 6.73 | 13.1 | 44.42 |
| No. UT. Acad. for Math Engineering & Science | NA | NA | NA |
| Noah Webster Academy | 4.18 | NA | 8.00 |
| North Davis Preparatory Academy | 6.57 | 12.64 | 7.32 |
| North Star Academy | 4.83 | 10 | NA |
| Odyssey Charter School | 9.26 | NA | 0.93 |
| Ogden Preparatory Academy | 2.40 | 2.03 | NA |
| Open Classroom | 11.69 | 11.11 | NA |
| Pacific Heritage Academy | 2.66 | 7.46 | 36.49 |
| Paradigm High School | NA | 35.09 | 19.12 |
| Pinnacle Canyon Academy | 17.88 | 13.25 | 4.65 |
| Pioneer High School for the Performing Arts | NA | NA | NA |
| Promontory School of Expeditionary Learning | 5.88 | 4.49 | 4.50 |
| Providence Hall | 6.16 | 5.24 | 17.07 |
| Quest Academy | 3.21 | 7.02 | NA |
| Ranches Academy | 5.50 | NA | NA |
| Reagan Academy | 2.29 | 4.55 | 22.14 |
| Real Salt Lake Academy High School | NA | NA | NA |
| Renaissance Academy | NA | NA | 22.58 |
| Rockwell Charter High School | NA | 14.63 | 66.67 |
| Roots Charter High School | NA | NA | NA |
| Salt Lake Arts Academy | 4.55 | 4.23 | 9.57 |

(Table continues)

**Table A3 continued**

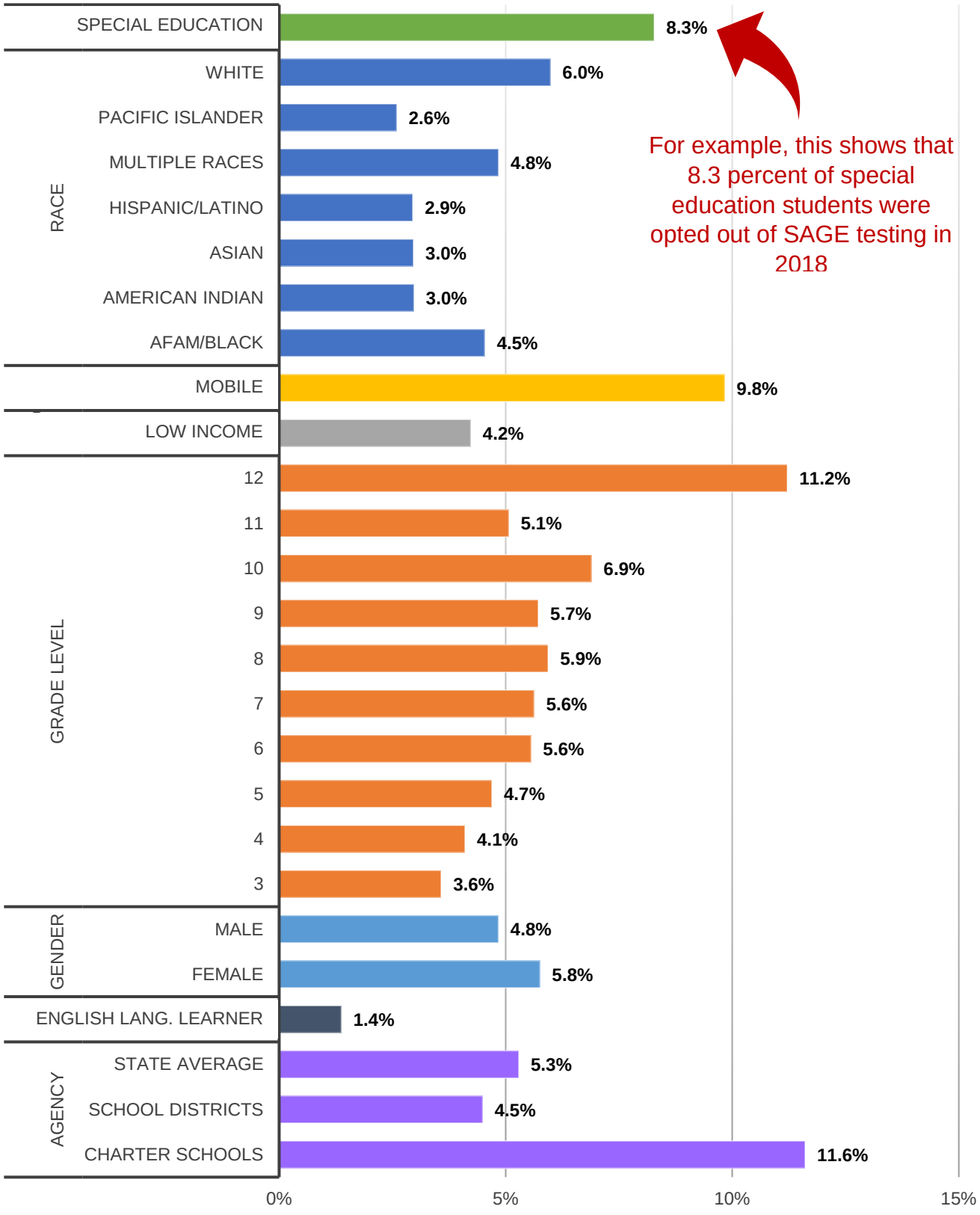
| School Name | Grades 3-6 | Grades 7-8 | Grades 9-11 |
|---|---|---|---|
| Salt Lake Center for Science Education | 4.29 | 0.91 | 11.66 |
| Salt Lake School for the Performing Arts | NA | NA | NA |
| Scholar Academy | 8.49 | NA | NA |
| Soldier Hollow Charter School | 9.09 | 0.00 | 31.25 |
| Spectrum Academy | 32.52 | 35.62 | 13.75 |
| St. George Academy | NA | 16.18 | 12.32 |
| Success Academy | NA | NA | NA |
| Summit Academy | 4.90 | 6.18 | 5.88 |
| Syracuse Arts Academy | 4.87 | 5.74 | 9.52 |
| Terra Academy | 8.21 | 16.33 | NA |
| The Center for Creativity, Innovation and Discovery | 19.35 | NA | 4.00 |
| Thomas Edison | 8.70 | 10.00 | NA |
| Timpanogos Academy | 72.32 | NA | NA |
| Treeside Charter School | 8.97 | NA | 20.47 |
| Tuacahn High School for the Performing Arts | NA | NA | 54.84 |
| Uintah River High | NA | NA | 25.00 |
| Utah Career Path High School | NA | NA | 6.92 |
| Utah County Academy of Science | NA | NA | 5.26 |
| Utah International Charter School | NA | 7.02 | 15.85 |
| Utah Military Academy | NA | 15.28 | NA |
| Valley Academy | 12.66 | 12.00 | 1.08 |
| Vanguard Academy | NA | 0.00 | 12.28 |
| Venture Academy | 4.83 | 8.85 | NA |
| Vista at Entrada School of Performing Arts and Technology | 10.89 | 9.18 | NA |
| Voyage Academy | 4.56 | NA | 24.24 |
| Walden School of Liberal Arts | 12.30 | 15.38 | NA |
| Wallace Stegner Academy | 3.98 | 2.35 | NA |
| Wasatch Peak Academy | 5.81 | NA | NA |
| Wasatch Waldorf Charter School | 16.84 | 20.75 | NA |
| Weilenmann School of Discovery | 5.57 | 17.36 | 8.62 |
| Winter Sports School | NA | NA | NA |

**This Page Left Blank Intentionally**

# Appendix B
# Overview of 2018
# Opt-Out Rates by Category

**This Page Left Blank Intentionally**

## Overview of 2018
## Opt-Out Rates by Category



**This Page Left Blank Intentionally**

# Appendix C
# Summary of Opt-Out Provisions in Other States

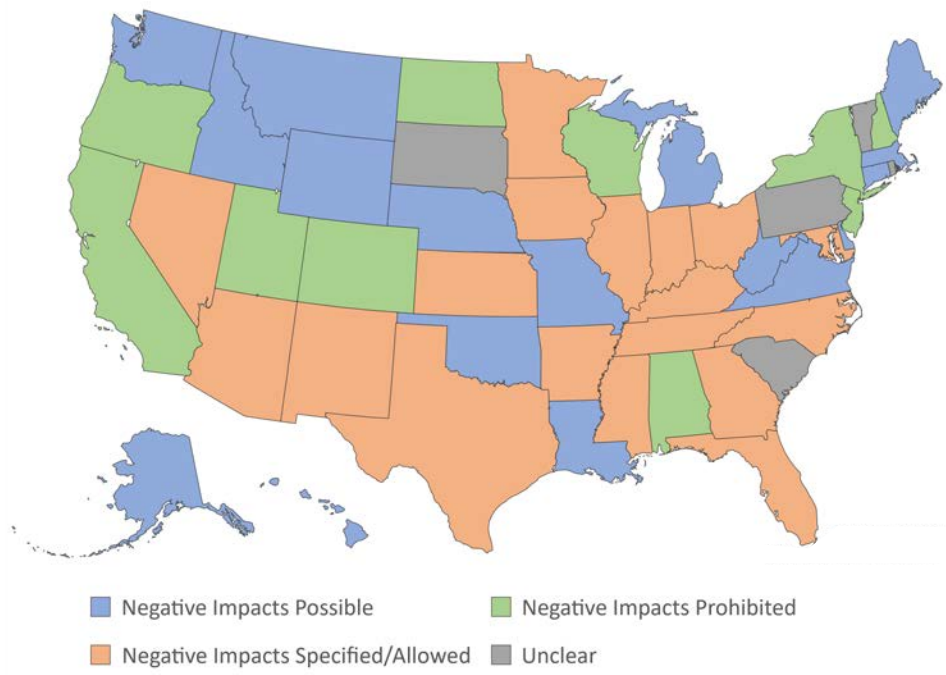
**This Page Left Blank Intentionally**

## Opt-Out Provisions Vary Widely Among States

To better understand how Utah's approach to opt-outs compares nationally, we researched opt-out provisions for other states. We contacted states by email and phone to inquire about their respective opt-out provisions and conducted online searches for information in cases where states did not respond to our inquiries.

We found states' approaches to addressing opt-outs are nuanced, making comparisons and categorizations challenging; therefore, our review focused on whether states specify and/or expressly allow negative impacts for students who do not participate in state assessments. Negative impacts might include preventing students from moving on to the next grade or from graduating. Figure C.1 summarizes our findings.

**Figure C.1 Summary of Other States' Opt-Out Provisions.** Only 10 states expressly prohibit consequences for students who are opted out of statewide assessments. Eighteen states specify or expressly allow consequences, and consequences are possible or unclear in nearly half of the states.




*Source: Auditor analysis of opt-out provisions in other states*

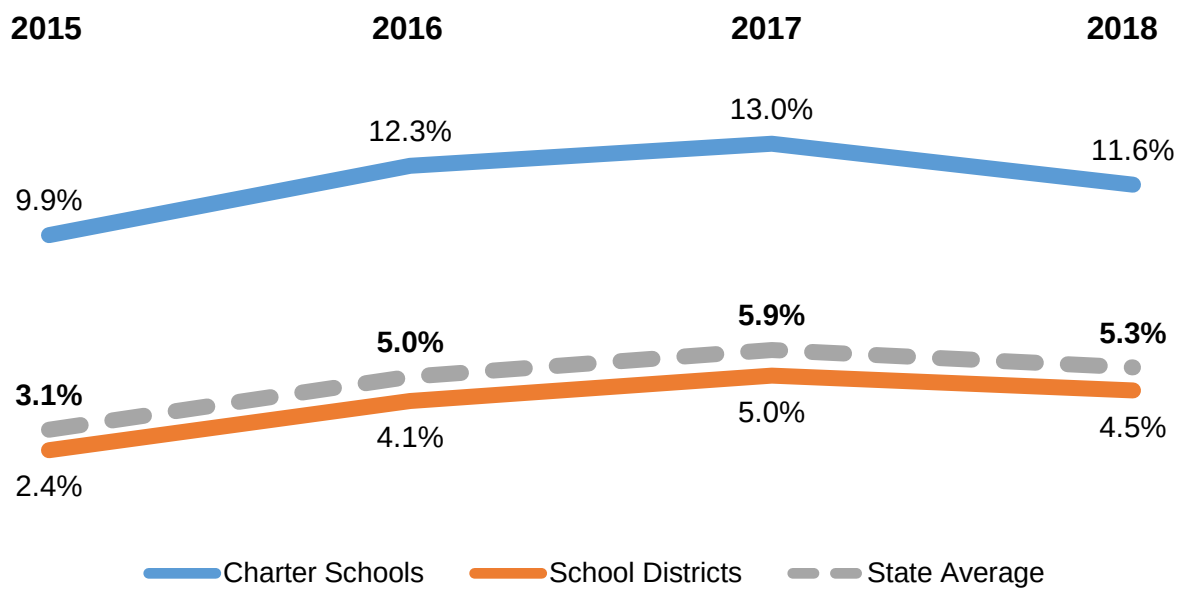
While many states are silent on the specific issue of opt-outs, Figure C.1 shows that negative impacts are possible or alluded to (but not specific or expressly allowed) in 18 states. For example, Missouri requires student participation, but does not specify consequences or allowance for consequences. In 5 states, the risk of consequences for students is unclear. For example, Vermont requires student participation but "does not act in any way on parent opt out requests."

Utah is one of only 10 states that specifically protect opted-out students from negative impacts. Utah statute provides a clear method for parents to opt their children out of statewide assessments in ***Utah Code* 53G-6-803(9)(a)**:

> At the request of a student's parent or guardian, an LEA shall excuse a student from taking an assessment that: (i) is federally mandated; (ii) is mandated by the state under this public education code; or (iii) requires the use of: (A) a state assessment system; or (B) software that is provided or paid for by the state.

***Utah Code* 53G-6-803** and ***Administrative Code* R277-404-7** authorizes parents to exempt (opt out) their children from state required assessments and provides protections against negative impacts for parents, students, LEAs, and LEA employees due to parental opt-out. Figure C.2 summarizes Utah's opt-out rates over time for school districts and charter schools.

**Figure C.2 Summary of 2015-2018 Opt-Outs.** This figure shows statewide opt-out rate averages for students in school districts and charter schools.




*Source: Auditor analysis of parental exclusion data provided by USBE*

Opt-out rates nearly doubled from 2015 to 2017 but decreased in 2018. While charter schools have much higher opt-out rates than school districts, the state average is usually less than 1 percent higher than the statewide school district rate. USBE also tracks opt-out data including demographical information such as race, gender, and low economic status. **Appendix B** contains a summary of this information for 2018.

In a letter to the U.S. Department of Education, the Utah State Superintendent of Public Instruction requested a waiver from certain federal requirements for student

participation in state required assessments and stated that "Utah policymakers strongly support parents' rights in directing and overseeing a student's education."

**This Page Left Blank Intentionally**

# Agency Response

**This Page Left Blank Intentionally**

November 19, 2019

Kade Minchey, CIA, CFE
Auditor General
Office of the Legislative Auditor General
W315 State Capitol Complex
Salt Lake City, UT 84114

Dear Mr. Minchey:

Thank you for the opportunity to respond to Audit Report 2019-12 "A Performance Audit of Public Education Assessment Data." The Utah State Board of Education (USBE) recognizes audits as a tool to identify and assess risks and improve the public education system. We appreciate the professionalism and courtesy of the audit staff in conducting the audit.

We concur with the recommendations to determine feasibility in collecting item response data, including time-per-item data, and evaluate the prevalence of test-taker disengagement. These actions as well as the recommendation to integrate the disengagement metrics into statewide standards assessment results, if determined to be feasible by the USBE, will require the USBE to work with the existing vendor (possibly a contract amendment) and our Technical Advisory Committee (TAC) to determine the most appropriate way to use the item response data in reporting. As such we request consideration in implementing the recommendations.

The information obtained through the consultant study is both interesting and informative. As the researchers included recommendations at the conclusion of their study, we are providing responses to those recommendations as well.

Recommendation 1: The USBE has a testing ethics policy that addresses these concerns. Each educator is required to annually complete a training on testing ethics and then sign a document, verifying they completed the training as provided by their LEA. The testing ethics policy includes explicit requirements related to educators' responsibilities in the administration of state tests.

Recommendation 2: The USBE publishes an Accountability Technical Manual that articulates uniform rules related to state assessment management, including the assignment of test participation codes. In addition, assessment directors and test administrators are provided annual

250 East 500 South  P.O. Box 144200  Salt Lake City, UT  84114-4200  Phone: (801) 538-7500

training on the process for entering participation codes. Likewise, the test delivery software allows administrators (LEA and SEA) to monitor and ensure participation codes are being entered for each test event.

Recommendation 3: If it is deemed important to conduct additional investigation regarding student motivation, we suggest it is appropriate to have the Office of the Legislative Auditor General conduct and report the research. If findings determine that a lack of student motivation is interfering with the validity of inference, any meaningful change to address the root cause would require legislative action.
Recommendation 4: While conducting additional research and future studies, to determine if student disengagement on state assessments is interfering with validity inferences could provide interesting and informative information, we don't believe there is sufficient evidence of its existence to justify the use of Utah taxpayer funds to complete the work.

Sincerely,

Patty Norman

Patty Norman
Utah State Board of Education
Deputy Superintendent of Student Achievement

cc: Mark Huntsman, Utah State Board of Education (USBE), Chair
Sydnee Dickson, USBE, State Superintendent of Public Instruction
Scott Jones, USBE, Deputy Superintendent of Operations
Darin Nielsen, USBE, Assistant Superintendent of Student Achievement
Debbie Davis, USBE, Chief Audit Executive

250 East 500 South P.O. Box 144200 Salt Lake City, UT 84114-4200 Phone: (801) 538-7500