



UNITED STATES DEPARTMENT OF EDUCATION
OFFICE OF INSPECTOR GENERAL

AUDIT SERVICES

Control Number
ED-OIG/A07M0001

March 31, 2014

James H. Shelton III
Deputy Secretary
U.S. Department of Education
400 Maryland Avenue, SW
Room 7W310
Washington, DC 20202

Dear Mr. Shelton:

This final audit report, “The U.S. Department of Education’s and Five State Educational Agencies’ Systems of Internal Control Over Statewide Test Results,” presents the results of our assessment of selected aspects of the systems of internal control over statewide test results designed and implemented by the U.S. Department of Education (Department), the Michigan Department of Education (Michigan), the Mississippi Department of Education (Mississippi), the Nebraska Department of Education (Nebraska), the South Carolina Department of Education (South Carolina), and the Texas Education Agency (Texas).

The objective of our audit was to determine whether the Department and the five State educational agencies (SEAs) had systems of internal control that prevented, detected, and required corrective action if they found indicators of inaccurate, unreliable, or incomplete statewide test results. Our audit of four of the five SEAs covered statewide tests administered during school years 2007–2008 through 2009–2010. However, because Nebraska used locally developed tests instead of statewide tests until school year 2009–2010, our audit of Nebraska covered statewide tests administered during school years 2009–2010 through 2011–2012.

We found that the Department and all five SEAs had systems of internal control designed to prevent and detect inaccurate, unreliable, or incomplete statewide test results. However, these systems did not always require corrective action if indicators of inaccurate, unreliable, or incomplete statewide test results were found. Furthermore, steps could be taken to improve the effectiveness of these systems. The Department could improve its monitoring of States’ test results by requiring SEAs to provide an explanation for data that the Department’s data collection system flagged as either incorrect or outside an anticipated range. It also could improve its monitoring of SEAs and local educational agencies (LEAs) by resuming reviews of test administration procedures during onsite monitoring visits and having SEAs’ systems of

internal control over statewide test results evaluated during standards and assessment peer reviews.

SEAs could improve their systems of internal control by (1) incorporating forensic analyses into their risk assessments to more effectively identify LEAs and schools with possible test administration irregularities, (2) strengthening their monitoring of LEAs' and schools' administration of statewide tests, (3) improving follow-up and resolution of test administration irregularities to prevent them from happening in the future, and (4) strengthening test security environments and test administration practices put in place by LEAs and schools.¹ The Department could help SEAs improve their systems of internal control by emphasizing, during its reviews of SEAs, the importance of using forensic analyses to more effectively identify schools with possible test administration irregularities.

We provided a draft of this report to the Department for comment. The Department stated that it agreed with our findings and all but two of our recommendations. Although the Department disagreed with draft report Recommendation 2.1, it proposed updating its standards and assessment peer review process to address the finding. The Department partially agreed with draft report Recommendation 2.2, stating that while forensic analyses can be useful, some forms of forensic analysis might no longer be applicable given changes to computerized testing. Therefore, the Department suggested the recommendation not be specific to the three types of forensic analyses discussed in the report. We summarized the Department's comments at the end of each finding and included the full text of its comments as Attachment 2 of this report.

We agree with the Department that updating its standards and assessment peer review process is a good idea and added a recommendation for the Department to include in its updated peer review manual procedures to review SEAs' systems of internal control over preventing, detecting, and requiring corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results (final report Recommendation 2.1). However, we did not revise or eliminate draft report Recommendation 2.1 (now Recommendation 2.2) because the Department does not annually monitor the test administration procedures implemented by every SEA. We believe it is important to obtain annual assurances that test administration procedures are operating as intended. We revised draft report Recommendation 2.2 (now Recommendation 2.3) to clarify that the types of forensic analyses used do not have to be the specific types discussed in this report. Finally, based in part on the Department's comments, we revised portions of the report for clarity.

¹ We issued separate audit reports to Michigan on May 20, 2013 (control number A07M0007), and Texas on September 26, 2013 (control number A05N0006), to address specific internal control weaknesses in those States.

BACKGROUND

Title I of the Elementary and Secondary Education Act of 1965, as amended (ESEA), requires States to adopt challenging academic content standards that specify what students are expected to know and be able to do. States also must adopt challenging student academic achievement standards that are aligned with the academic content standards. To measure the success of students in meeting the academic achievement standards, States must establish a set of high-quality, yearly student academic tests. The tests must measure the proficiency of all students in, at a minimum, mathematics, reading or language arts, and science. The tests also must be valid, reliable, and consistent with relevant, nationally recognized professional and technical standards.

States must establish a single minimum percentage of students who are required to meet or exceed the proficient level on these tests. According to Section 1111(b)(1)(D)(ii)(II) of the ESEA, States must describe two levels of high achievement, proficient and advanced, that determine how well students are mastering academic material in the State academic content standards. States also must describe a third level of achievement, basic, to provide complete information about the progress of the lower-achieving students toward meeting the proficient and advanced levels of achievement. Student performance on the yearly student academic tests is then used as the primary means in determining the yearly performance of the State, each LEA, and each public school in the State.

In 2011, the Department began offering States flexibility from certain ESEA requirements. According to Department guidance, “ESEA Flexibility,” issued September 23, 2011, and updated June 7, 2012, an SEA may request, on its own behalf and on behalf of its LEAs, flexibility through waivers of 10 provisions of the ESEA and their associated regulatory, administrative, and reporting requirements. Although an SEA may receive waivers related to determining adequate yearly progress, to receive a waiver, an SEA must have developed, or have a plan to develop, annual, statewide, high-quality tests and corresponding academic achievement standards that measure student progress towards mastering academic material for grades three through eight and at least once in high school. The Department considers a high-quality test to be a test or a system of tests that is valid, reliable, and fair for its intended purposes and measures student knowledge and skills against college- and career-ready standards in reading or language arts and mathematics.

As of November 2013, the Department had approved the waiver requests made by 42 SEAs, the District of Columbia, and Puerto Rico. The Department still was reviewing waiver requests from three SEAs and the Bureau of Indian Education. Five SEAs have not submitted waiver requests. Of the five SEAs that we selected as part of this audit, Michigan, Mississippi, South Carolina, and Texas had their waiver requests approved by the Department. Nebraska has not submitted a waiver request.

Within the Department, the Office of Elementary and Secondary Education has a mission to promote academic excellence, enhance educational opportunities and equity for all of America’s

children and families, and improve the quality of teaching and learning by providing leadership, technical assistance, and financial support. Within the Office of Elementary and Secondary Education, the Office of Student Achievement and School Accountability Programs (SASA) is responsible for ensuring that States develop the academic standards and student assessment and accountability systems that are needed to hold LEAs and schools accountable for the academic progress of their students.

SASA reviews and recommends approval or disapproval of amendments to State accountability plans. It also usually coordinates peer review teams that review the adequacy of assessment systems when States propose changes to those systems. According to the 2009 peer review manual, reviewers were to evaluate whether the State established clear criteria for the administration, scoring, analysis, and reporting components of its assessment system and whether the State had a system for monitoring and improving the ongoing quality of its assessment system. The review manual did not address evaluating specific systems of internal control over statewide test administration and security.

In December 2012, the Department temporarily suspended peer reviews of State assessment systems. According to a letter that the Office of Elementary and Secondary Education issued to chief State school officers on December 21, 2012, the Department's decision was based on two considerations. One, the Department wanted to permit States to focus their resources on designing and implementing new tests that will provide a better measure of critical thinking skills and complex student learning to support good teaching and improved student outcomes—and not on their old systems. Two, the Department wanted an opportunity to review the peer review process and criteria to determine what changes might improve the process. In the letter, the Department noted that it would consider enhancing aspects of the peer review process, including a State's test security policies and procedures.

SASA also monitors States' uses of Title I, Part A, and other ESEA program funds via onsite monitoring and desk reviews. SASA developed indicators to determine the extent that States have implemented the programs and activities that the States administer. SASA's monitoring typically covers three areas: standards, assessment, and accountability; instructional support; and fiduciary responsibilities.

Alleged Cheating on Statewide Tests

Allegations of cheating on statewide tests have been reported in multiple States and the District of Columbia. Our analysis of media reports on cheating that occurred during the past 10 years showed that the most prevalent methods of cheating included the following:

1. using actual test questions to prepare students for the tests,
2. erasing students' wrong answers and filling in the correct answers,
3. indicating the correct answers to students during testing,

4. allowing students to change answers after giving them the correct answers, and
5. allowing students to discuss answers with each other.

Other alleged methods of cheating included (1) completing incomplete test booklets, (2) altering attendance records, (3) failing to cover testing materials during the exams, (4) arranging the classroom to facilitate cheating, (5) reading questions aloud to students who were not eligible for that accommodation, (6) not testing all eligible students, and (7) obtaining testing materials when not authorized to do so. Actions that SEAs and LEAs have taken in response to substantiated allegations of cheating included (1) formally reprimanding or firing teachers and those involved, (2) hiring outside firms to investigate what happened, (3) having students retake tests, and (4) adding security measures, such as outside monitors to watch as tests were administered.

Cheating could make the data that are used to determine State-, LEA-, and school-level achievement under ESEA unreliable. Improper determinations made using unreliable school achievement data potentially affect Federal funding, school turnaround programs, and additional services for the most disadvantaged students.

Selection of States, LEAs, and Schools

As part of our methodology for selecting States, LEAs, and schools, we analyzed statewide test results data. States upload required test results data to the Department's Education Data Exchange Network (EDEN) Submission System, part of the Department's *EDFacts* initiative. *EDFacts* is a Department initiative that centralized K–12 performance data supplied by States with other data, such as financial grant information within the Department, to enable better analysis and use in policy development, planning, and management. *EDFacts* includes data on student proficiency on statewide tests, participation rates on tests, and graduation rates at the State, LEA, and school levels. States upload aggregated student test results to *EDFacts*. Our analysis of *EDFacts* data estimated probabilities of seeing observed year-to-year changes in grade and subject-level proficiencies within a school based on statewide patterns.²

Based on this analysis and other factors, we selected five States to review: Michigan, Mississippi, Nebraska, South Carolina, and Texas. From each State, we selected three LEAs. From each LEA, we selected one or two schools (see Table 1 on pages 5–6). We selected the LEAs because they had schools with relatively high year-to-year test score fluctuations. We also considered the number of schools in the LEAs, the average adjusted gross income (AGI) levels in the areas where the LEAs are located, and the LEA's adequate yearly progress trends (see Objective, Scope, and Methodology).

² We conducted our analysis to select States, LEAs, and schools for internal control reviews, not to determine whether cheating occurred.

Table 1. SEAs, LEAs, and Schools Selected

LEAs and Schools Visited	School Year 2009–2010 Student Enrollment	School Year 2009–2010 Title I, Part A Allocation to the LEA
Michigan—663 LEAs with at least one school with more than 200 students		
Cesar Chavez Academy <ul style="list-style-type: none"> • Cesar Chavez Academy Middle School 	1,601	\$ 2,589,746
Detroit Public Schools <ul style="list-style-type: none"> • John R. King Academic and Performing Arts Academy • William Beckham Academy 	81,151	\$312,243,605
School District of the City of Inkster <ul style="list-style-type: none"> • Baylor Woodson Elementary School • Blanchette Middle School 	2,559	\$ 3,565,019
Mississippi—138 LEAs with at least one school with more than 200 students		
George County School District <ul style="list-style-type: none"> • Central Elementary School • Agricola Elementary School 	4,237	\$1,552,013
Holly Springs School District <ul style="list-style-type: none"> • Holly Springs Junior High School • Holly Springs High School 	1,186	\$1,484,926
Rankin County School District <ul style="list-style-type: none"> • Richland High School • Pelahatchie Attendance Center 	15,189	\$3,600,208
Nebraska—112 LEAs with at least one school with more than 200 students		
Central City Public Schools <ul style="list-style-type: none"> • Central City Elementary School 	779	\$ 170,724
Columbus Public Schools <ul style="list-style-type: none"> • Centennial Elementary School • Emerson Elementary School 	3,632	\$ 703,913
Lincoln Public Schools <ul style="list-style-type: none"> • C. Culler Middle School • Belmont Elementary School 	32,784	\$ 13,427,442
South Carolina—84 LEAs with at least one school with more than 200 students		
Charleston County School District <ul style="list-style-type: none"> • Chicora School of Communications • Mt. Zion Elementary School 	37,037	\$32,765,870
Darlington County School District <ul style="list-style-type: none"> • St. John's Elementary School • Washington Street Elementary 	9,041	\$ 7,236,644
Lancaster County School District <ul style="list-style-type: none"> • Buford Elementary School • Erwin Elementary School 	10,928	\$ 6,230,836

LEAs and Schools Visited	School Year 2009–2010 Student Enrollment	School Year 2009–2010 Title I, Part A Allocation to the LEA
Texas—935 LEAs with at least one school with more than 200 students		
La Joya Independent School District <ul style="list-style-type: none"> • Sam Fordyce Elementary • La Joya High School 	26,401	\$31,826,479
Lufkin Independent School District <ul style="list-style-type: none"> • Brandon Elementary School • Lufkin High School 	5,878	\$ 4,731,192
Marion Independent School District <ul style="list-style-type: none"> • Marion Middle School • Marion High School 	1,053	\$ 220,062

AUDIT RESULTS

The objective of our audit was to determine whether the Department and the five SEAs had systems of internal control that prevented, detected, and required corrective action if they found indicators of inaccurate, unreliable, or incomplete statewide test results. Our audit of four of the five SEAs covered statewide tests administered during school years 2007–2008 through 2009–2010. However, because Nebraska used locally developed tests instead of statewide tests until school year 2009–2010, our audit of Nebraska covered statewide tests administered during school years 2009–2010 through 2011–2012.

We found that the Department and all five SEAs had systems of internal control designed to prevent and detect inaccurate, unreliable, or incomplete statewide test results. However, these systems did not always require corrective action if indicators of inaccurate, unreliable, or incomplete statewide test results were found. Furthermore, we identified steps could be taken to improve the effectiveness of these systems.

- The Department could improve its monitoring of States’ test results by requiring SEAs to provide an explanation for data that the *EDFacts* system flagged as either incorrect or outside an anticipated range. It could also improve its monitoring of SEAs and LEAs by resuming reviews of test administration procedures during its Title I program monitoring visits and having SEAs’ systems of internal control over statewide test results evaluated during standards and assessment peer reviews.
- SEAs could improve their systems of internal control by (1) incorporating forensic analyses into their risk assessments to more effectively identify LEAs and schools with possible test administration irregularities, (2) improving monitoring of LEAs’ and schools’ administration of statewide tests, (3) improving follow-up and resolution

of test administration irregularities to prevent them from happening in the future, and (4) strengthening test security environments and test administration practices put in place by LEAs and schools. The Department could help SEAs improve their systems of internal control by issuing updated and additional guidance that highlights promising practices in these areas.

In its response to the draft of this report, the Department stated that it agreed with our findings and all but two of our recommendations. Although the Department disagreed with draft report Recommendation 2.1 (now Recommendation 2.2), it proposed updating its standards and assessment peer review process to address the finding. Also, the Department partially agreed with draft report Recommendation 2.2 (now Recommendation 2.3), stating that while forensic analyses can be useful, some forms of forensic analysis might no longer be applicable given changes to computerized testing. Therefore, the Department suggested the recommendation not be specific to the three types of forensic analyses discussed in the report. We summarized the Department's comments at the end of each finding and included the full text of its comments as Attachment 2 of this report.

We agree with the Department that updating its standards and assessment peer review process is a good idea and added a recommendation for the Department to include in its updated peer review manual procedures to review SEAs' systems of internal control over preventing, detecting, and requiring corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results (final report Recommendation 2.1). However, we did not revise or eliminate draft report Recommendation 2.1 (now Recommendation 2.2) because the Department does not annually monitor the test administration procedures implemented by every SEA. We believe it is important to obtain annual assurances that test administration procedures are operating as intended. We revised draft report Recommendation 2.2 (now Recommendation 2.3) to clarify that the types of forensic analyses used do not have to be the specific types discussed in this report. Finally, based in part on the Department's comments, we revised portions of the report for clarity.

FINDING NO. 1 – The Department Could Strengthen Its Monitoring of States' Test Results and Test Administration Procedures

The Department monitored test results and test administration procedures by using data validation checks on data that SEAs submitted to the Department through *EDFacts* and following up on flagged results. To minimize risks related to the collection and review of statewide test results, the Department's data collection system is available only to a limited number of users, password protected, and open only for certain periods. The Department also (1) requires SEAs to have *EDFacts* coordinators; (2) encourages the use of State-level manuals on test administration, accommodations, and security; (3) provides technical assistance to SEAs when they are developing their data collection systems; and (4) provides training to *EDFacts* coordinators. The Department also developed rules and edit checks that alert SEAs about potential problems with data as the data are being entered. However, the Department has not always required SEAs to provide explanations for every flagged result.

The Department also reviewed test administration procedures during its onsite monitoring visits to SEAs. The Department visits SEAs and LEAs on a 3-year cycle. SASA developed indicators to determine the extent to which States have implemented Federal programs and activities that the SEAs administered. Monitoring covered three areas: standards, assessment, and accountability; instructional support; and fiduciary responsibilities. The monitoring of standards, assessment, and accountability included a review of SEA and LEA policies and procedures for test security, data quality, training, and monitoring test administration procedures. However, the Department suspended its reviews of test administration procedures during monitoring visits in school year 2011–2012 because of other programmatic priorities.

The Department Did Not Always Require SEAs to Provide an Explanation for Data That the ED*Facts* System Flagged as Either Incorrect or Outside Anticipated Ranges

SEAs submit data on student proficiency on statewide tests, participation rates, and graduation rates to the Department through ED*Facts*. These data from ED*Facts* are used to create the Consolidated State Performance Report (CSPR). The Department used the CSPR to monitor an SEA's progress in implementing 15 ESEA programs, including Title I, Part A. The Department also used the CSPR to identify areas in which the SEA might need technical and program management assistance and areas in which policy changes might be needed. For school year 2011–2012, the Department required SEAs to certify that, to the best of their knowledge, CSPR data were true, reliable, and valid.

The Department also created automated checks for data entered into ED*Facts* and used to create the CSPR. The purpose of these automated checks was to identify missing data or possible problems with statewide test results. The automated checks flagged data that were either incorrect or outside anticipated ranges.

For State-level data used to create the CSPR, the Department defined three types of flags: (1) fatal flaw, (2) flag with comment, and (3) flag without comment. The system rejected all data flagged as fatal flaws until SEAs corrected the data. The system required the SEA to provide an explanation before it would accept data that were flagged with comment. However, for a flag without comment, the system accepted the data without requiring the SEA to provide an explanation for why the data were either incorrect or outside an anticipated range.

One scenario when the system would flag data without comment is when the percentage of students who tested at or above the proficient level increased or decreased by at least 15 percentage points from the previous year. The system flagged without comment the State of Georgia statewide test data for school years 2008–2009 and 2009–2010 for this reason.³ However, the Department did not require an explanation. Had the Department required an explanation for why the State of Georgia's test data fell outside the anticipated ranges, indicators of potential cheating on statewide tests in the State of Georgia might have been detected sooner.

³ The Department does not review the data at the LEA and school levels.

The Department Suspended Its Reviews of Test Administration Procedures During Its Title I Program Monitoring at SEAs and LEAs

The Department conducts Title I program monitoring visits at every SEA and selected LEAs every 3 years. During school years 2007–2008 through 2010–2011, SASA reviewed SEA and LEA policies and procedures for test security, data quality, training, and monitoring of test administration as part of its Title I program monitoring. However, SASA suspended reviews of test administration procedures during school year 2011–2012 program monitoring visits. According to SASA, it suspended these reviews because it planned to focus on the waivers that States were granted under ESEA flexibility.

SASA informed us that it began reviewing test administration procedures again in April 2013. During its reviews of Montana, Nebraska, and Vermont, SASA used an updated monitoring plan that included questions focused on training, securing tests, identifying test irregularities, and responding to test irregularities. By including these questions in the program monitoring plan, SASA has strengthened its monitoring protocols for test environments at SEAs and LEAs. As of May 2013, SASA had not decided whether it would review test administration procedures during all its onsite monitoring visits.

According to “Standards for Internal Control in the Federal Government,” November 1999,⁴ monitoring is one of the five standards of internal control. Monitoring should assess the quality of performance over time and ensure that the findings of audits and other reviews are promptly resolved. Sound monitoring processes should include evaluations focused on the effectiveness of the controls. In the case of ensuring the effectiveness of controls over the validity of statewide test results, we believe it is necessary for SASA to monitor test administration procedures at all SEAs.

Recommendations

We recommend that the Assistant Secretary for Elementary and Secondary Education—

- 1.1 Require SEAs to provide explanations for all statewide test data that are used to create the CSPRs and that *EDFacts* classifies as flagged without comment.
- 1.2 Ensure that SASA reviews, using its updated monitoring plan, test administration procedures and data quality at all SEAs.

Department Comments

The Department agreed with the finding and recommendations. The Department stated that SASA continually reviews all flags and follows up with either a letter or phone call. The Department also reiterated that State officials receive all CSPR flags before certifying the final CSPR submission data are accurate. The Department also stated that SASA continually updates its monitoring plans and has developed additional monitoring question for testing data quality. The additional monitoring questions are based on prior Office of Inspector General and Government Accountability Office reports regarding test security, as well as National Council of

⁴ The United States Government Accountability Office.

Measurement of Education professional standards. The Department stated it had used the updated monitoring plan in its reviews of Montana, Nebraska, and Vermont.

Office of Inspector General Response

Based on the Department's comments, we clarified in the finding that, in some instances, the Department required SEAs to provide explanations for every flagged result. We also clarified in the finding that, in addition to Nebraska, the Department has completed reviews of Montana and Vermont using its updated monitoring plans. We did not make any other substantive changes to this finding.

FINDING NO. 2 – SEAs Could Strengthen Their Oversight of Statewide Test Administration and Security

In “Key Policy Letters from the Education Secretary or Deputy Secretary,” June 24, 2011, the Department urged chief State school officers to review and, if necessary, strengthen their States’ efforts to protect student achievement and accountability data, ensure the quality of those data, and enforce test security. Despite the Department’s guidance, not all SEAs have taken the opportunity to better protect student achievement and accountability data. We found that SEAs could strengthen their oversight of statewide test administration and security by (1) incorporating forensic analyses into their risk assessments to more effectively identify LEAs and schools with possible test administration irregularities, (2) improving their monitoring of LEAs’ and schools’ administration of statewide tests, (3) improving follow-up and resolution of test administration irregularities to prevent them from happening again, and (4) strengthening test security environments and test administration practices put in place by LEAs and schools. The Department could help SEAs strengthen their oversight of test administration and security by updating and adding guidance on promising practices in these four areas.

Forensic Analyses Needed to More Effectively Identify LEAs and Schools With Possible Test Administration Irregularities

All five SEAs that we reviewed monitored schools for possible test administration irregularities by conducting onsite monitoring visits at LEAs and schools or following up on irregularities that LEAs reported to the SEA. However, four of these five SEAs either did not incorporate or incorporated only limited forensic analyses in their risk assessment and monitoring procedures. These SEAs could improve their ability to detect possible test administration irregularities by making use of more forensic analyses in their risk assessment and monitoring procedures.

According to “Testing Integrity Symposium Issues and Recommendations for Best Practice,” published by the Department in February 2013, there are three primary approaches for detecting testing irregularities using forensic analysis: test-score analysis, ratio analysis and erasure analysis, and item-response pattern analysis. Test-score analysis examines test scores for unusually large gains from the previous year or large declines in the subsequent year. Test-score analysis can be used to look at student-level data as well as classroom- or school-level results. Ratio analysis and erasure analysis evaluates the number of answers that were changed from a wrong to a right response. This analysis allows an SEA to identify patterns that warrant further investigation. Item-response pattern analysis can identify unusually common response patterns across students within the same class. SEAs that use computer-based testing will not be able to

use erasure analyses but could make use of other types of analyses, such as response-time and sequence-of-response analysis, that are not available in a paper-based testing environment.

Four of the five SEAs that we visited used at least one type of forensic analysis, annually spending as little as \$12,700 and as much as \$128,000 for forensic analysis. One SEA used test-score analysis and erasure analysis to identify schools to monitor for possible test administration irregularities. A second SEA used erasure analysis to identify schools to monitor. A third SEA did not use forensic analysis to identify LEAs or schools to monitor but did use forensic analysis to follow up on possible test administration irregularities that LEAs reported to it. A fourth SEA did not use forensic analysis to identify schools for monitoring during our audit period but started using forensic analysis subsequent to our audit period. The fifth SEA did not use forensic analysis.

Although the third of these five SEAs did not use test-score or erasure analysis to identify LEAs or schools for monitoring, it did write its contract in such a way that its statewide testing contractor would provide erasure analysis if the SEA deemed it necessary to gather more information during reviews of possible test administration irregularities. If the SEA asked its statewide testing contractor to perform erasure analysis, the contractor would analyze classroom-level data. The SEA considered a wrong-to-right erasure count to be excessive if the average number of wrong-to-right erasures for all students in a classroom exceeded the State average by three standard deviations.

We identified a weakness to using the classroom-level approach. If a classroom had a small number of individuals with anomalous numbers of erasures, then averaging these erasures for all students in a classroom could hide these individual anomalies. For example, if a classroom had 20 students, and 1 student had 20 wrong-to-right erasures while the other 19 students had no wrong-to-right erasures, the average would be only 1 wrong-to-right erasure per student. A reviewer would see only that the classroom had a low number of average wrong-to-right erasures per student and not notice that one student had a large number of wrong-to-right erasures. Furthermore, we found that one of the schools that we visited reported about 400 students in one grade as one classroom. Combining this many students into one classroom would make it difficult for the analysis to show a classroom-level anomaly.

As noted, during our audit period, the fourth of the five SEAs did not use erasure analysis alone to identify schools for further monitoring or to identify test administration irregularities, and the fifth SEA did not use test-score or erasure analysis at all, albeit for different reasons. Although the fourth SEA considered a number of wrong-to-right erasures that exceeded the State average by four standard deviations to be excessive, it did not use this data and follow up with schools that had students with an excessive number of wrong-to-right erasures. In school year 2011–2012, this SEA began using erasure analysis to identify schools to follow up with for possible test administration irregularities. It also changed its criteria so that it considered a count of erasures of any type that exceeded two standard deviations above the State average to be excessive.

The fifth SEA did not use test-score or erasure analysis to identify schools for further monitoring because SEA officials believed that they did not have enough statewide test score data available

to review. Prior to school year 2009–2010, LEAs in the State formulated, administered, and scored the tests but sent only aggregated data rather than student-level data to the State. Beginning in school year 2012–2013, the reading, math, and science tests were computerized. Therefore, erasure data were not available to analyze.

Forensic analyses can be effective tools to use during the risk assessment process and could provide SEAs with more comprehensive risk identification. For example, using erasure analyses and analyzing proficiency rates, we found that two of the five schools in one State that we reviewed had one class with students who had an excessive number of wrong-to-right erasures at the same time that the classes had unusually large gains in year-to-year proficiency rates.⁵ Using a statistical analysis,⁶ we determined that at least 20 percent of the students in these classes had a number of wrong-to-right erasures with a less than 1 in 10,000 chance of occurring at random based on statewide test results. These two schools also had anomalous changes in student proficiency that had a probability of occurrence of less than 1 in 10,000. There are a variety of possible explanations for anomalous wrong-to-right erasure counts and changes in proficiency, and sanctions should not be based solely on forensic analysis. However, anomalous results on two different analyses in the same classroom at the same school demonstrate a level of risk that warrants follow-up.

If SEAs do not use at least one type of forensic analysis, they might not identify and investigate potential indicators of test administration irregularities that could lead to inaccurate, unreliable, or incomplete test results. This lack of monitoring might result in LEAs and schools repeatedly violating test administration and security procedures, leading to an increased risk of inaccurate and unreliable statewide test results and a missed opportunity to detect and prevent cheating.

Onsite Monitoring During the Administration of Tests and Sharing the Results Can Be an Effective Tool for Ensuring the Validity of Test Results

SEAs and LEAs conduct onsite monitoring visits at schools to verify that schools follow appropriate test administration procedures. Onsite monitoring consisted of activities such as observing test administration, verifying that test materials were properly secured, and verifying that schools provided required accommodations. Four of the five SEAs conducted onsite monitoring of LEAs and schools during test administration and documented the results of their monitoring. The fifth SEA did the same through school year 2010–2011. After school year 2010–2011, however, the SEA trained LEAs to conduct their own onsite monitoring of schools.

Four of the five SEAs that we visited also shared the results of the monitoring visits directly with school principals or LEA officials or presented issues uncovered during onsite monitoring in training sessions. The fifth SEA could improve its monitoring of test administration by providing the results of its onsite monitoring visits to the LEAs and schools and monitoring schools that it identifies as high risk of having test administration irregularities. Although this SEA conducted onsite monitoring visits during test administration, it did not always share the results of the visits with the LEAs and schools.

Three SEAs did not use risk factors, such as the number and types of test administration irregularities that occurred in prior years, to select LEAs and schools for onsite monitoring visits.

⁵ We brought the results of our analyses to the SEA's attention, and it agreed to follow up with the two schools.

⁶ For details on the statistical analysis used, see the Objectives, Scope, and Methodology section of this report.

One SEA identified eight LEAs in 2010 that had reported intentional violations of test administration standards that compromised the validity of the tests. However, the SEA did not conduct onsite monitoring visits at any of those LEAs in the following year. Another SEA did not use a risk assessment process to select LEAs for onsite monitoring. Instead, it selected LEAs geographically to ensure that it covered the areas of each State board of education member. The third SEA relied on LEAs to conduct their own onsite monitoring even if test administration irregularities were reported in prior years.

SEAs and LEAs should conduct onsite monitoring visits at schools to verify that schools follow appropriate test administration procedures. According to 34 C.F.R. § 80.40(a), grantees are responsible for monitoring grant- and subgrant-supported activities to ensure compliance with applicable Federal requirements and the achievement of performance goals.

In “Key Policy Letters from the Education Secretary or Deputy Secretary,” June 24, 2011, the Department encouraged SEAs to review and, if necessary, strengthen efforts to protect student achievement and accountability data, ensure the quality of those data, and enforce test security. The Department suggested conducting unannounced onsite visits during test administration.

If an SEA does not monitor schools that it thinks might have test administration irregularities, then the schools might never identify the irregularities and correct them before test results are compromised. If an SEA does not share its monitoring results with the LEAs and schools, then the LEAs and schools will not be aware of potential weaknesses that they should correct.

SEAs Could Improve Their Follow-Up and Resolution of Test Administration Irregularities to Prevent Them From Happening in the Future

SEAs could improve their handling of test administration irregularities by documenting corrective action recommendations and the resolution of irregularities and creating a process that ensures that LEAs and schools timely report all irregularities.

All five SEAs that we reviewed imposed sanctions for test administration irregularities, encouraged LEAs and schools to report test administration irregularities, and provided information on how to report improper conduct. As a result, LEAs generally reported test security breaches to the SEAs. Following the SEAs’ guidance, the LEAs also informed their employees of the expectations of conduct and performance related to administering statewide tests and provided information on how to report improper conduct.

Despite these efforts, two of the five SEAs that we visited could improve their handling of test administration irregularities to prevent future incidents. One SEA did not timely resolve potential test administration irregularities. We identified 38 cases from school years 2009–2010 through 2011–2012 that were closed. The number of lapsed days between when these cases were reported to the SEA and when the cases were officially closed ranged from 291 to 517 days. At the time of our analysis, an additional five cases still were open and had been pending for 520 to 941 days. Rather than assigning a single employee the role of updating the database that tracks test administration irregularities, the SEA distributed the responsibility among multiple employees; therefore, follow-up on pending cases might not have occurred in a timely manner.

Another SEA did not always document the actions that it required LEAs to take to correct test administration irregularities or whether the LEAs implemented the corrective actions. In school year 2009–2010, one of the three LEAs that we reviewed had three irregularities for which it could not document that it took any corrective action: 13 students were not tested, 2 students completed answer documents assigned to other students, and 1 student received an incorrect test score because of a miscoded answer document. The SEA did not maintain documentation identifying the corrective action that it required the LEA to take or documentation showing that the LEA took any corrective action. In addition, the SEA did not verify that the LEA took corrective action.

In addition to the weaknesses that we identified at these two SEAs, 1 of 15 LEAs that we reviewed did not completely or timely report test administration irregularities to the SEA. During school years 2007–2008 through 2009–2010, the LEA did not report 53 irregularities to the SEA. The unreported irregularities included giving students the wrong tests or answer sheets, misplacing test booklets and answer sheets, and leaving students unattended during testing. During school years 2009–2010 through 2011–2012, the LEA reported 13 irregularities to the SEA more than 6 months after the irregularities occurred. The LEA also could not determine how many reviews it conducted during school year 2007–2008 because it did not retain documentation for all reviews. It retained documentation for only nine reviews that showed test administration irregularities during the year.

In “Key Policy Letters from the Education Secretary or Deputy Secretary,” the Department urged chief State school officers to review and, if necessary, strengthen efforts to protect student achievement and accountability data, ensure the quality of those data, and enforce test security. The Department suggested that chief State school officers seek support to enact strict and meaningful sanctions against individuals who transgress the law or compromise professional standards of conduct.

According to “Testing Integrity Symposium, Issues and Recommendations for Best Practice,” a clear and comprehensive test security policy should include strong and clear language addressing protocol for reporting breaches and sanctions for misconduct. If an investigation concludes that cheating has occurred, SEAs and LEAs should be clear about what sanctions that SEAs may impose.⁷

If SEAs do not have effective processes for timely resolving test administration irregularities, documenting their corrective action recommendations and resolutions, and ensuring that LEAs and schools timely report all irregularities, SEAs and LEAs might not resolve irregularities in a consistent and appropriate manner in time to correct them before the next test administration. Further, by not taking action when irregularities are discovered, an appropriate message is not conveyed about the proper administration of tests. This can compromise the integrity of test results.

⁷ For guidelines on test administration and security developed by States through the Council of Chief State School Officers, see “TILSA Test Security Guidebook: Preventing, Detecting, and Investigating Test Security Irregularities,” John F. Olson and John Fremer, May 2013, available at: http://ccsso.org/Resources/Publications/TILSA_Test_Security_Guidebook.html.

In “Key Policy Letters from the Education Secretary or Deputy Secretary,” the Department states that just the hint of irregularities or misconduct during test administration can call into question school reform efforts and compromise the State’s accountability system. According to “Testing and Data Integrity in the Administration of Statewide Student Assessment Programs,” October 2012,⁸ test administration irregularities compromise the reliability and validity of the test results, which causes the public to lose confidence in the testing program and in the educational system.

SEAs Could Enhance Test Security Environments and Test Administration Practices

All five SEAs that we reviewed had procedures in place to promote secure test administration environments. Each SEA had a written code of conduct that described the ethical requirements for its employees. Although these codes did not have requirements that were specific to test administration, they promoted general ethical practices that would enhance test security environments.

All five SEAs also required personnel involved with test administration (for example, district and school test coordinators and school test administrators) to sign agreements stating that they would protect the security and integrity of statewide tests. These agreements included some or all of the following assertions: the employee was trained in test security, read and understood testing procedures, and understood the consequences of not complying with testing procedures. Four SEAs required all test administrators to sign these agreements. The fifth SEA did not require test administrators to sign the agreements but did require test coordinators and school principals to sign the agreements.

Weaknesses in test security environments

Although all five SEAs that we reviewed had procedures in place to promote secure test administration environments, we identified weaknesses in providing secure test environments at two SEAs and three LEAs. One SEA stored student information in a database that was used to create labels for assigning test materials to students. Although the SEA required users to have a password to access the database, it did not require users to regularly change their passwords. Access security control protects the systems and network from inappropriate access and unauthorized use. An example of a specific control activity is frequent changes of passwords. If users do not periodically change their passwords, unauthorized users who acquire the passwords might access the system. The SEA could strengthen its data security by requiring users to change their passwords periodically.

Another SEA required its testing contractor to provide LEAs and schools with missing test materials reports. These reports showed the scorable (answer documents) and nonscorable (for example, test booklets) test materials that LEAs and schools never returned to the testing contractor. One of the LEAs that we reviewed had missing nonscorable materials. However, the contractor did not provide the report to the SEA until approximately 7 months after the LEA administered the tests. If the contractor does not timely provide the report, then those in possession of the materials have an extended time to review and evaluate the materials,

⁸ National Council on Measurement in Education.

compromising test security. The SEA could strengthen its test security by requiring its contractors to timely report missing test materials.

According to “Testing Integrity Symposium, Issues and Recommendations for Best Practice,” LEAs should establish a clear chain of custody and retain control over test administration materials to prevent tampering. Test administrators should limit access to materials, record when and where the materials were accessed, and clearly communicate that there are consequences for failing to secure the materials.

All five SEAs had different document numbers assigned to every test, and three SEAs had procedures requiring all tests to be sealed. In addition, most of the LEAs and schools that we reviewed secured test materials in locked rooms before and after test administration, limited access to these rooms to those who needed materials to administer tests, and required those receiving the materials to sign them in and out. They also followed SEA guidance for administering tests. However, we identified the following weaknesses in LEAs’ and schools’ procedures for securing test materials:

- One LEA’s building security allowed for unauthorized access to test materials.
- Two schools at different LEAs allowed teachers to keep test materials in their classrooms overnight during test administration, even though SEA policy prohibited such a practice. At one of the schools, test administrators were not trained in securing test materials.

Weaknesses in test administration procedures

We also identified the following weaknesses in procedures for test administration at the LEA and school levels:

- One LEA administered the wrong tests to students, incorrectly coded information on answer documents, and provided answer documents to the wrong students. These irregularities caused incorrect test results.
- One school allowed high school teachers to administer subject tests to their own students even though the LEA recommended that teachers not be allowed to administer tests to their own students. Teachers have a stake in the test results of their own students; therefore, they might have an incentive to manipulate test results. Administering the tests to their own students might give them the opportunity to do so.
- One school did not report to the LEA that it did not test students in a continuous session. It allowed students who could not finish a test before lunch to break for lunch with other students who did not finish the test and resume testing after lunch. Taking breaks during a test gives students the opportunity to discuss the test with other students before they complete the test.

According to “Testing Integrity Symposium, Issues and Recommendations for Best Practice,” tests should be administered in controlled and secure environments. To reduce the risk that test administration irregularities will occur, SEAs, LEAs, and schools must ensure that test administration environments are secure and that LEAs and schools follow appropriate test security and administration procedures.

Department Guidance

In 2003 and 2006, the Department issued guidance for improving data quality in States’ and LEAs’ annual report cards that provide information on program effectiveness, academic results, and teacher quality. In nonregulatory guidance issued on September 12, 2003, the Department reminded States and LEAs to refer to the data quality guidelines posted on the Department’s Web site. In “Improving Data Quality for Title I Standards, Assessments, and Accountability Reporting, Guidelines for States, LEAs, and Schools, Non-Regulatory Guidance,” April 2006, the Department presented a set of guidelines to address data quality issues associated with the annual report cards required of all States, LEAs, and schools receiving Title I funds. The guidelines provided information regarding good practices in data collection and were intended to enhance internal control over data used to make key judgments about adequate yearly progress, funding, accountability, and other State and LEA policies. However, this guidance did not discuss test security and administration practices.

The Department began issuing guidance on test security and administration practices to States in 2011. In “Key Policy Letters from the Education Secretary or Deputy Secretary,” the Department urged chief State school officers to make test security a high priority. The Department urged chief State school officers to review and, if necessary, strengthen efforts to protect student achievement and accountability data, ensure the quality of those data, and enforce test security. Suggested efforts included

- conducting a risk assessment of LEA- and school-level capacity to implement test security and data quality procedures;
- ensuring that test development contracts include support for activities related to monitoring test administration, including forensic analyses;
- conducting unannounced onsite visits during test administration to review compliance with professional standards on test security; and
- seeking support to enact strict and meaningful sanctions against individuals who transgress the law or compromise professional standards of conduct.

The Department also gave a presentation, “Prevention and Detection of Test Irregularities,” to the National Association of Title I Directors on February 10, 2012. The presentation covered policies and procedures for

- preventing test administration irregularities from occurring;
- training test administrators and monitoring test administration;

- detecting test administration irregularities, including using forensic analyses; and
- following up on possible test administration irregularities.

In February 2013, the Department published “Testing Integrity Symposium Issues and Recommendations for Best Practice.” The report describes promising practices for

- preventing test administration irregularities,
- detecting and analyzing test administration irregularities,
- responding to and investigating possible test administration irregularities, and
- ensuring the integrity of online and technology-based tests.

The guidance that the Department has provided during the past 3 years should provide SEAs, LEAs, and schools with tools that they need to prevent, detect, and require corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results. However, we believe that the Department could take action to help them even more by providing additional, specific examples of strong systems of internal control.

Recommendations

We recommend that the Assistant Secretary for Elementary and Secondary Education—

- 2.1 Update the standards and assessment peer review manual to include an evaluation of SEAs’ systems of internal control over preventing, detecting, and requiring corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results.
- 2.2 Require SEAs to annually certify to the Department that they have systems of internal control that prevent, detect, and require corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results.
- 2.3 Determine, during future Title I monitoring visits to SEAs, whether the SEAs have implemented appropriate methods, such as one of the types of forensic analyses described in “Testing Integrity Symposium Issues and Recommendations for Best Practice,” to identify schools with possible test administration irregularities.

We also recommend that the Assistant Secretary for Elementary and Secondary Education provide updated guidance to SEAs that encourages them to—

- 2.4 Monitor schools that they identify as high risk for having test administration irregularities and share the results of their monitoring with LEAs and schools.

- 2.5 Strengthen prevention and handling of test administration irregularities by (a) creating a formal process for timely resolving test administration irregularities, (b) documenting their corrective action recommendations and the resolution of irregularities, and (c) creating a process that ensures that LEAs and schools timely report all irregularities.
- 2.6 Strengthen test security environments by ensuring that (a) test administration systems are secure, including having users periodically change their passwords; (b) contractors timely provide all test administration reports to SEAs; (c) LEAs and schools store test materials in locked rooms at all times when students are not taking the tests; and (d) LEAs train test administrators in test security.
- 2.7 Strengthen test administration practices by ensuring that LEAs and schools put in place procedures that will prevent irregularities from occurring.

Department Comments

The Department agreed with the finding and draft report Recommendations 2.3 through 2.6 (now Recommendations 2.4 through 2.7). The Department disagreed with draft report Recommendation 2.1 (now Recommendation 2.2), questioning whether annual certifications would produce the desired result—assurance of an effective of system of internal control. The Department suggested that the recommendation focus on addressing this issue through the Department’s updated standards and assessment peer review process.

The Department partially agreed with draft report Recommendation 2.2 (now Recommendation 2.3), noting that some forms of forensic analyses might no longer be applicable given that States are moving to computerized testing. The Department agreed that forensic analyses can be useful but suggested that the recommendation be modified to focus more generally on SEAs having implemented appropriate measures to identify testing irregularities, with forensic analyses provided as one example.

Office of Inspector General Response

We agree with the Department that updating its standards and assessment peer review process is a good idea and added a recommendation for the Department to update the manual to include an evaluation of SEAs’ systems of internal control over preventing, detecting, and requiring corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results (final report Recommendation 2.1). However, we did not revise or eliminate draft report Recommendation 2.1 (now Recommendation 2.2) because the Department does not annually monitor the test administration procedures implemented by every SEA. We believe it is important to obtain annual assurances that the test administration procedures are operating as intended. Accordingly, we believe that requiring SEAs to annually certify that they have systems of internal control that prevent, detect, and require corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results is the best way to obtain such assurances.

We revised draft report Recommendation 2.2 (now Recommendation 2.3) to clarify that the types of forensic analyses used do not have to be specific to the types discussed in the report.

We also added to the finding examples of forensic analyses in a computerized testing environment.

OBJECTIVE, SCOPE, AND METHODOLOGY

The objective of our audit was to determine whether the Department and the five SEAs had systems of internal control that prevented, detected, and required corrective action if they found indicators of inaccurate, unreliable, or incomplete statewide test results. Our audit of four of the five SEAs covered statewide tests administered during school years 2007–2008 through 2009–2010. However, because Nebraska used locally developed tests instead of statewide tests until school year 2009–2010, our audit for Nebraska covered statewide tests administered during school years 2009–2010 through 2011–2012.

To achieve our objective, we did the following:

- Obtained background information about the programs, activities, and entities being audited.
- Reviewed and gained an understanding of
 1. sections 1111, 1112, 1116, 1117, and 6111 of the ESEA, as amended;
 2. regulations at 34 C.F.R. §§ 80.40, 80.42, and 200;
 3. “Standards for Internal Control in the Federal Government,” November 1999;
 4. “Key Policy Letters from the Education Secretary or Deputy Secretary,” June 24, 2011, and “Testing Integrity Symposium Issues and Recommendations for Best Practice,” February 2013; and
 5. “Internal Control-Integrated Framework,” September 1992.⁹
- Reviewed prior Office of Management and Budget Circular A-133 compliance audits for the 5 selected SEAs and 15 selected LEAs and Department program monitoring reports on all 5 SEAs to identify areas of potential internal control weaknesses related to our audit objective.
- Reviewed written policies and procedures and testing records at the 5 SEAs, 15 LEAs, and 28 schools. At the SEAs, we also reviewed contracts for scoring tests.

⁹ The Committee of Sponsoring Organizations of the Treadway Commission.

- Interviewed officials from SASA and the Department's Office of Planning, Evaluation, and Policy Development; SEA officials responsible for statewide tests; and superintendents, district test coordinators, school test coordinators, school test administrators, and test proctors at the LEAs and schools.
- Gained an understanding and assessed the adequacy of the Department's system of internal control over preventing, detecting, and requiring corrective action if they found indicators of inaccurate, unreliable, or incomplete statewide test results by
 1. identifying and reviewing the controls the Department had for collecting and reviewing statewide assessment data;
 2. reviewing Title I program monitoring plans and the peer review assessment manual, and
 3. reviewing the guidance provided by the Department to SEAs regarding the integrity of assessment results and test security.
- Gained an understanding and assessed the adequacy of the systems of internal control at each of the 5 SEAs, 15 LEAs, and 28 schools by obtaining from each entity, when available, and reviewing
 1. procedures for administering statewide tests;
 2. procedures for reviewing test results to ensure the accuracy and validity of the results;
 3. monitoring procedures and checklists;
 4. procedures for identifying, reporting, and following up on test administration irregularities;
 5. sanctions imposed for test security violations; and
 6. guidance documents, such as administration manuals, test memoranda, emails, and handouts provided during training, that they provided on test administration, test security, and reporting of test administration irregularities.

We also analyzed erasure data for the 17 schools that we visited in Michigan, South Carolina, and Texas to identify whether any students had an excessive number of wrong-to-right erasures that might indicate that there was a test administration irregularity. Michigan and South Carolina provided erasure data for individual students, and Texas provided erasure data for classrooms. We analyzed the data and calculated the probability of seeing the observed wrong-to-right erasure count for each student or classroom. We calculated the probabilities using a generalized

Poisson distribution¹⁰ based on statewide wrong-to-right erasure data because we found that a generalized Poisson distribution fit the observed data. To detect anomalous counts, we flagged any wrong-to-right erasure count having less than a 1 in 10,000 chance of occurring at random given statewide test results. We did not analyze erasure data for schools in Mississippi because Mississippi did not maintain erasure data. Instead, it received reports from a third-party contractor that already had analyzed the erasure data. We reviewed the third-party contractor reports for the three LEAs that we visited and those LEAs did not have any anomalous wrong-to-right erasure counts. We did not analyze erasure data for schools in Nebraska because Nebraska used computerized tests.

Finally, we reviewed the comments on the draft of this report that we received from the Department on March 18, 2014, and revised the report accordingly.

Data Reliability

To achieve our objective, we relied on data from *EDFacts*. *EDFacts* included data on student proficiency on statewide tests, participation rates, and graduation rates at the State, LEA, and school levels. We used these data to select States, LEAs, and schools to visit as part of this audit.

To determine whether the *EDFacts* data for the five SEAs were accurate and complete, we reconciled proficiency scores for selected grades and subjects at selected schools that we calculated using data that each SEA provided us with the scores that we obtained from *EDFacts*. We performed this reconciliation of data from each of the five SEAs for school years 2007–2008, 2008–2009, and 2009–2010. We did not find any discrepancies between the scores recorded in *EDFacts* and the scores that we calculated using data for school years 2007–2008 through 2009–2010 that we obtained from Michigan, Mississippi, Nebraska, and Texas. Although South Carolina did not retain documentation to support the data that it submitted to the Department for school year 2007–2008, we applied selected logic tests, such as looking for missing data and reviewing the relationship of one data element to another, to that year's *EDFacts* data for South Carolina. We did not find any anomalies in the school year 2007–2008 data for South Carolina. We also did not find any discrepancies between the student proficiency scores recorded in *EDFacts* and the scores that we calculated using data for school years 2008–2009 and 2009–2010 that we obtained from South Carolina. Therefore, we concluded that the data from *EDFacts* for school years 2007–2008 through 2009–2010 for all five States were sufficiently reliable for our intended use.

We also relied on erasure data that the SEAs provided for school years 2008–2009 and 2009–2010 for the 17 schools that we visited in Michigan, South Carolina, and Texas. To determine whether the erasure data for the 17 schools that we visited in Michigan, South Carolina, and Texas were reliable, we gained an understanding of each SEA's processes for reviewing the contractors' scoring procedures. We also performed logic tests on the data. We looked for missing data and the relationship of one data element to another. Based on our understanding of each SEA's processes and the results of our tests, we concluded that the data were sufficiently reliable for our intended use.

¹⁰ A Poisson distribution is a discrete frequency distribution that gives the probability of a number of independent events occurring in a fixed time.

Sampling Methodology

We judgmentally selected 5 States, 15 LEAs, and 28 schools for review.

Selection of States

We used *EdFacts* data to identify all schools in 47 States and the District of Columbia with total enrollment of more than 200 students during school years 2007–2008, 2008–2009, and 2009–2010.¹¹ For each school with a total enrollment of more than 200 students, we calculated risk scores for each grade tested in the subjects of math and reading. We calculated a risk score to determine how anomalous each increase or decrease in proficiency for a grade and subject was from one year to the next in relation to the change for that grade or subject across each respective State.

After calculating a series of risk scores for each school, we used the maximum risk score from each school. We calculated the average maximum risk score by State to identify 33 States that had an average maximum risk score that was near or below the national average. We eliminated the other 14 States from consideration because the States with average school maximum risk scores that were above the national average might have had large score fluctuations because of broad changes in curricula or tests. In the 33 States, there were 8,798 LEAs. We looked for LEAs with highly unlikely score patterns and identified 70 LEAs that had at least one school with very high-risk scores. We considered any observed maximum risk score with a less than 1 in 10,000 chance of occurring at random given statewide test results to be anomalous. We selected Michigan as our first State because it had three LEAs with a school that had multiple grades with anomalous risk scores. We used the risk score only to assist with the selection of States, LEAs, and schools for internal control reviews, not to determine whether cheating occurred at a particular LEA or school.

After conducting work in Michigan, we revised two aspects of our selection methodology before selecting the additional four States. First, we realized that a composite risk score, rather than a maximum risk score, tended to detect anomalies across multiple grades. The revised methods calculated a school composite risk score, which was the average of the school's five highest risk scores across all years, grades, and subjects. Second, we classified LEAs by Internal Revenue Service-reported AGI because we found that our initial methodology tended to focus only on lower income areas. To find the average AGI for the LEA, we calculated a student count weighted average based on the AGI for each school's zip code. We considered only the 6,197 LEAs that had 3 or more schools. We then excluded 4 of the 6,197 LEAs from our universe because we did not have average AGI data for the areas that those LEAs served. Therefore, our universe consisted of 6,193 LEA.

For the final State selections, we used *EDFacts* data for school years 2007–2008, 2008–2009, and 2009–2010 to calculate the composite risk score for each school based on the school's five highest risk scores. In addition, we assigned all LEAs having at least three schools to one of three groups based on average AGI (see Table 2).

¹¹ We excluded New Jersey, Vermont, and Wyoming because, for these three States, *EDFacts* did not have the data that we needed for our analysis.

Table 2. LEA Category Definitions Based on Average AGI

LEA AGI Category	Range of Average AGI	Number of LEAs
Highest Third	\$52,675 or greater	2,066
Middle Third	\$41,900 to \$52,674	2,067
Lowest Third	Less than \$41,900	2,060

We selected Mississippi, Nebraska, and South Carolina because they had the three highest average composite risk scores. We selected Texas because it had 19 LEAs with average composite risk scores in the top 5 percent of LEAs nationally and the fourth largest standard deviation in the average composite risk scores of its LEAs. Furthermore, all 4 of these States had at least 1 LEA in each AGI category with an average composite risk score in the top 5 percent of LEAs nationally.

Selection of LEAs and Schools

We identified the following universes of LEAs in each selected State that had at least 1 school with more than 200 students during school years 2007–2008, 2008–2009, and 2009–2010:

- Michigan: 663
- Mississippi: 138
- Nebraska: 112
- South Carolina: 84
- Texas: 935

From those universes, we selected three LEAs in each State (see Table 1 on pages 5–6 of this report). In Michigan, we selected these LEAs because they were the only three LEAs that had schools with multiple grades with maximum risk scores greater than 10. In the other four States, we selected the LEAs because they had high average composite risk scores. We also considered the number of schools in the LEA and the LEA’s history of adequate yearly progress status. We selected LEAs from each of the average AGI categories.

We judgmentally selected six schools in Mississippi, South Carolina, and Texas, two from each of the selected LEAs. We selected five schools in Michigan and Nebraska, at least one school from each of the selected LEAs.

In Michigan, we selected schools that had multiple grades with maximum risk scores that corresponded to a less than 1 in 10,000 chance of occurring at random given statewide results. In the other four States, we selected schools that had high average composite risk scores, high fluctuations in math and reading proficiency rates, and annual changes in whether they met adequate yearly progress requirements. See Table 1 on pages 5–6 of this report for a list of the selected schools.

We conducted this audit from December 2011 through August 2013 in

- Detroit, Inkster, and Lansing, Michigan;
- Brandon, Holly Springs, Jackson, and Lucedale, Mississippi;
- Central City, Columbus, and Lincoln, Nebraska;
- Charleston, Columbia, Darlington, and Lancaster, South Carolina;
- Austin, La Joya, Lufkin, and Marion, Texas; and
- Washington, D.C.

We discussed the results of our audit with Michigan officials on September 5, 2012, and January 8, 2013; Mississippi officials on September 17, 2013; Nebraska officials on September 4, 2013; South Carolina officials on August 30, 2013; Texas officials on May 29, 2013; and Department officials on September 19, 2013. We issued a separate audit report to Michigan on May 20, 2013. We issued a separate audit report to Texas on September 26, 2013.

We conducted this performance audit in accordance with generally accepted government auditing standards (July 2007 and December 2011 revisions). Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions on our audit objective. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions on our audit objective.

ADMINISTRATIVE MATTERS

Corrective actions proposed (resolution phase) and implemented (closure phase) by your office will be monitored and tracked through the Department's Audit Accountability and Resolution Tracking System (AARTS). Department policy requires that you develop a final Corrective Action Plan (CAP) for our review in AARTS within 30 days of the issuance of this report. The CAP should set forth the specific action items and targeted completion dates necessary to implement final corrective actions on the findings and recommendations contained in this final audit report.

In accordance with the Inspector General Act of 1978, as amended, the Office of Inspector General is required to report to Congress twice a year on the audits that remain unresolved after 6 months from the date of issuance.

In accordance with the Freedom of Information Act (5 U.S.C. § 552), reports issued by the Office of Inspector General are available to members of the press and general public to the extent information contained therein is not subject to exemptions in the Act.

We appreciate the cooperation given us during this audit. If you have any questions, please call me at (202) 245-6900 or Gary D. Whitman, Regional Inspector General for Audit, at (312) 730-1620.

Sincerely,

/s/

Patrick J. Howard
Assistant Inspector General for Audit

ATTACHMENT 1

Acronyms, Abbreviations, and Short Forms Used in this Report

AGI	Adjusted Gross Income
CSPR	Consolidated State Performance Report
Department	U.S. Department of Education
EDEN	Education Data Exchange Network
ESEA	Elementary and Secondary Education Act of 1965, as amended
LEA	Local Educational Agency
Michigan	Michigan Department of Education
Mississippi	Mississippi Department of Education
Nebraska	Nebraska Department of Education
SASA	Office of Student Achievement and School Accountability Programs
SEA	State Educational Agency
South Carolina	South Carolina Department of Education
Texas	Texas Education Agency

ATTACHMENT 2



UNITED STATES DEPARTMENT OF EDUCATION
OFFICE OF THE DEPUTY SECRETARY

March 10, 2014

MEMORANDUM

TO: Patrick J. Howard
Assistant Inspector General for Audit
Office of Inspector General

FROM: James H. Shelton, III Acting Deputy Secretary
U.S. Department of Education

SUBJECT: Response to the Draft Audit Report “The U.S. Department of Education’s and Five State Educational Agencies’ Systems of Internal Control Over Statewide Test Results” (Control No. ED-OIG/A07M0001)

Thank you for the opportunity to review the Office of Inspector General’s (OIG) draft audit report titled “U.S Department of Education’s and Five State Educational Agencies’ Systems of Internal Control Over Statewide Test Results” (ED-OIG/A07M0001).

The Department is strongly committed to helping ensure that there are reliable and valid measures of student performance. We appreciate OIG’s work on this important subject and found the report to be balanced and informative. Many in the Department who participated in the review and discussions with OIG staff about the importance of having rigorous internal controls in statewide testing results welcomed the opportunity to discuss the challenges State and local officials face in managing the fast-paced changes and innovations in assessments. The dynamic landscape of statewide testing that includes the increasing need for robust measures of complex student learning outcomes, the need to measure both performance and growth, the use of extended responses and accommodations, the use of spiraling forms, the innovative use of technology in assessments, and the desire to assess students in more personalized learning environments, will require more innovative and frequently updated ways to maintain strong internal control systems.

Test security policies and procedures must be reviewed and updated to address these innovative uses of technology in testing and assessment. Now, more than ever, it is critical that institutions are able to demonstrate that their students are developing the essential skills needed for success.

We welcome the opportunity to work with state and local officials to help them improve test security policies and procedures that should be reviewed and updated frequently to address next generation assessments. Your report will help in these ongoing efforts.

The Office of Elementary and Secondary Education (OESE) agrees with the findings and, in part, with the recommendations. Specifically, OESE agrees with recommendations 1.1, 1.2, 2.3, 2.4, 2.5, 2.6 and with recommendations 2.1 and 2.2 as modified in the attached. Below you will find some summary comments on the report. Additionally, we have marked specific comments and suggestions for revisions in the attached copy of the audit report.

Response to the Draft Audit Report “The U.S. Department of Education’s and Five State Educational Agencies’ Systems of Internal Control Over Statewide Test Results” (Control No. ED-OIG/A07M0001)

The Office of Student Achievement and School Accountability relies to some degree on an important set of professional guidelines developed by States through the Council of Chief State School Officers (CCSSO): TILSA Test Security Guidebook: Preventing, Detecting, and Investigating Test Security Irregularities.¹ This CCSSO guide was prepared by CCSSO’s Technical Issues in Large Scale Assessment (TILSA) group engaged with experts in assessments. The guidebook provides key standards for states to prevent, detect, and investigate test security irregularities as assessment and accountability undergo substantial changes. Practical implementation samples address test security standards, recommended provisions for test security manuals, examples of data forensics information for requests for proposals, sample security investigations kit, model language for state and district-level academic dishonesty policies, and many more resources for SEA and LEA self-analysis, on-site monitoring, and forensic analysis. The State guidebook provides extensive resources to States on key issues raised by the OIG in its draft report, and OIG may wish to reference this in the audit report as an additional source of information to help State and local governments with these issues.

¹ CCSSO: TILSA Test Security Guidebook: Preventing, Detecting, and Investigating Test Security Irregularities, by John F. Olson & John Fremer. (May, 2013). Available at: http://ccsso.org/Resources/Publications/TILSA_Test_Security_Guidebook.html

**Specific comments provided by the Department on sections of the draft audit report.
These comments were in addition to the two-page letter.**

Comment 1 (page 2 of the report)

We suggest providing some additional background on ESEA flexibility. We have included some suggested edits. In 2011, as part of an initiative known as “ESEA flexibility,” the Department began offering States flexibility from certain ESEA requirements in exchange for implementing rigorous, comprehensive State-developed plans designed to improve educational outcomes for all students, close achievement gaps, increase equity, and improve the quality of instructions. Additionally, you may want to consider including information on the number of entities that have applied for and have been approved for flexibility, such as the following:

As of November 2013, 45 states, the District of Columbia, Puerto Rico and the Bureau of Indian Education submitted requests for ESEA flexibility, and 42 States, the District of Columbia and Puerto Rico have been approved for ESEA flexibility. The waivers were generally approved for two years, and applicants can request an extension of the waiver. Additionally, a coalition of districts in California applied for and received waivers of certain ESEA requirements in exchange for locally developed plans to prepare all students for college and career, focus aid on the neediest students, and support effective teaching and leadership.

Comment 2 (page 3 of the report)

We suggest clarifying this language to maintain consistency with the Department’s ESEA flexibility guidance. We have included the suggested edits:

Although an SEA may receive waivers related to determining adequate yearly progress, to receive this flexibility an SEA must have developed, or have a plan to develop, annual, statewide, high-quality tests and corresponding academic achievement standards in at least reading or language arts and mathematics for grades three through eight and at least once in high school that are aligned with college- and career- ready standards and that measure student growth.

Comment 3 (page 3 of the report)

Please note that Texas’ ESEA flexibility request was approved on September 30, 2013.

Comment 4 (page 3 of the report)

We suggest clarifying the reasons why peer review has been suspended to be more consistent with the letter in December 2012. We also suggest noting the explicit reference to test security procedures in the letter. We have included suggested edits:

However, the Department has temporarily suspended peer reviews of State assessment systems. According to a letter that the Office of Elementary and Secondary Education issued to chief State school officers on December 21, 2012, the Department’s decision was based on two considerations: To permit States to

focus their resources on designing and implementing new assessments that will provide a better measure of critical thinking skills and complex student learning to support good teaching and improved student outcomes—and not on their old systems—and to provide the Department an opportunity to review the current peer review process and criteria to determine what changes might improve that process, especially in light of the transition by most States to the next generation of assessments aligned to college- and career-ready standards. In the letter, the Department noted that it would consider enhancing aspects of the peer review process, including a State’s test security policies and procedures.

Comment 5 (page 4 of the report)

The list of the most prevalent methods of cheating from the results of OIG’s analysis of media reports on cheating that occurred during the past 10 years is informative. If available it would also be helpful to include a list of some of the state sanctions in place by state statute or regulation in the five states to address test cheating, such as suspension or revocation of teaching and administrator licenses.

Comment 6 (page 8 of the report)

As to explanations for all statewide test data that are used to create the CSPRs and that *EDFacts* classifies as “flagged without comment,” we ask that the OIG acknowledge ongoing improvements made in the review of *EDFacts* submissions. SASA staff continually reviews all flags and follows up with either a letter or phone call. It is important to note that state officials are provided all CSPR flags before they certify the final CSPR submission; thus by virtue of certification the State is ensuring that data submitted is accurate. SASA also participates in cross-office coordination throughout ED to ensure agency-wide data policy consistency. We have included suggested edits:

We acknowledge that the Department is engaged in ongoing improvements to its process for reviewing *EDFacts* submissions. Among these improvements, SASA staff continually reviews all flags, including flags without comment, and follows up with either a letter or phone call. States must address these issues before certifying their final CSPR submission as accurate. Additionally, SASA coordinates with other offices throughout the Department to ensure consistency with respect to data policy.

Comment 7 (page 9 of the report)

We suggest including additional language regarding SASA’s updated monitoring plans, including the development of additional monitoring questions regarding testing data quality. We have included suggested edits:

SASA noted that it continually updates its monitoring plans and has developed additional monitoring questions for testing data quality. These additional monitoring questions are based on prior OIG and Government Accountability Office (GAO) reports regarding test security, as well as National Council of Measurement in Education (NCME) professional standards. During its review of the States of Nebraska, Vermont, and Montana, SASA used an updated

monitoring plan that included questions focused on training, securing tests, identifying test irregularities, and responding to test irregularities.

Comment 8 on page 19 of report

While we agree that SEAs should have the type of systems of internal control referenced in this recommendation, we question the efficacy of addressing this through an annual certification. Instead, we would suggest reframing the recommendation to focus on addressing this issue through the Department's updated standards and assessment peer review process, and require amendments when there are significant changes. We have included suggested edits:

Include in its updated standards and assessment peer review procedures a review of SEA systems of internal control to prevent, detect, and require corrective action if they find indicators of inaccurate, unreliable, or incomplete statewide test results.

Comment 9 (page 19 of the report)

We have included the following suggested edit:

We recognize that many States are moving toward computer-based assessments and, as assessment technology evolves, erasure analysis and other more traditional forms of forensic analyses may no longer be relevant. However, computer-based assessment will allow for more sophisticated approaches using other types of data collected, such as response time and sequence of responses.

As noted above, given that some forms of forensic analysis may no longer be applicable given changes to technology related to assessments, we have some concerns regarding the emphasis on forensic analysis in this report. While we agree that such analyses can be useful, we would suggest that the recommendation be modified to focus more generally on SEAs having implemented appropriate measures to identify testing irregularities, with forensic analyses provided as one example. We have included suggested edits:

Determine, during future Title I monitoring visits to SEAs, whether the SEAs have implemented appropriate measures, such as one of the three types of forensic analyses described in "Testing Integrity Symposium Issues and Recommendations for Best Practice" to identify schools with possible test administration irregularities.