caveon™

Test Security

Test Security and Peer Review
Language for State Assessment RFPs

Caveon Test Security

February 2018

Test Security and Peer Review

Language for State Assessment RFPs

caveon™ Test Security

Introduction

Effective test security policies, procedures and practices are critical components of successful K-12 State accountability assessment programs and can help ensure the validity of assessment results and the important decisions made based upon those results. Security requirements must be clearly described in State Requests for Proposals (RFPs) or similar documents to ensure that intended security services are included in vendor bids for assessment programs. In addition, the US Department of Education requires test security validity evidence as part of federal Peer Review requirements.

This white paper provides valuable concepts and suggestions to State assessment professionals who write or review RFPs and describes the test security elements required not only to ensure the fairness, reliability and validity of test results, but also satisfy Peer Review test security requirements. Multiple components of test security, including the prevention, deterrence, and detection of test theft and cheating are all essential components of a strong test security plan that will support test validity and fair testing practices.

This white paper:

- Describes how test security practices and procedures can provide evidence of validity
- Identifies how important test security services can satisfy Peer Review evidence requirements
- Includes specific language that can be used directly in procurements to ensure vendors correctly interpret the intended test security guidelines in Appendix A

Test security should be explicitly addressed both in existing policies and procedures and in procurement language when testing programs are being created, renewed or redesigned. Various elements of test security complement each other to provide overall protection for a testing program. For example, results from data forensics analyses might inform the selection of schools for monitoring in successive administrations. Findings from internet monitoring can inform the design and development of different item types to enhance security. Investigations into irregularities may indicate areas where policies and procedures need to be developed or where additional training may be required.

The U.S. Department of Education Peer Review of State Assessment Systems Non-Regulatory Guidance for States[1] recognizes that comprehensive test security practices are part of a valid assessment program. The guidelines include two critical elements that intersect with test security practices:

- Critical Element 2.4 – Monitoring Test Administration
- Critical Element 2.5 – Test Security

These Critical Elements define the validity evidence that States should collect to ensure that test security procedures and practices are in place. The remaining section of this paper discuss the various elements of a test security program and how these elements relate to the required peer review evidence. In

[1] Education, U.S. Department of. Peer Review of State Assessment Non-Regulatory Guidance for States (PDF). Ed.gov.US Department of Education, 25 Sept. 2015.

addition, we describe what the test security component of a State assessment RFP should require of potential bidders. The following areas are covered:

- Independent Audit of Test Security Policies and Procedures
- Test Security Handbooks
- Data Forensics
- Web and Media Monitoring
- Test Administration Monitoring
- Investigations of Testing Irregularities

Additional procurement language is included in Appendix A. Appendix B provides an excerpt from the Dec. 8, 2015, Congressional Record including a colloquy allowing the use of Title I funding for test security services that improve the quality, validity, and reliability of State academic assessments.

Independent Audit of Test Security Policies and Procedures

States or other organizations that sponsor high-stakes assessments should periodically review their security policies and practices. These reviews, or audits, are most effective when conducted by external, objective test security experts.

A test security audit should include the process for evaluating the State Education Agency's (SEA's) test security policies, processes, and procedures against established standards. This evaluation should provide feedback to the SEA on areas of satisfactory or exemplary test security process and areas needing improvement. An initial presentation of finding should be provided to the SEA, followed by a written report of recommendations for improvement after conducting staff interviews, and extensive review of all manuals, descriptions, flowcharts, and allowable agreements.

Test Security Handbooks

Each SEA should create and maintain a handbook which documents the State's test security policies, procedures, and processes. The handbook will make use of the findings of the security audit and can provide:

- The detailed steps of how the SEA will administer tests in a secure environment
- The individuals who are responsible for test security and their roles and responsibilities
- How the SEA will respond in the case of a test security breach
- The tools it will use to prevent, deter, and detect/react to testing irregularities indicative of cheating

The security audit should result in recommendations for improvement as well as recognition of robust test security practices. Security audits can identify areas where testing practices need to be improved or where additional quality assurance practices should be implemented. Security handbooks can define how test security practices work together and intersect to support the validity of the testing program. For example, security policies on data forensics might indicate that forensics results should inform the

selection of schools for monitoring in successive administrations. These reviews can identify areas where additional quality assurance is needed or where training of testing personnel may be necessary.

Peer Review Validity Evidence – Test Security Handbooks

Critical Element 2.5 requires prevention, detection, remediation and investigation of testing irregularities. This Critical Element requires evidence to document "an appropriate set of policies and procedures to prevent test irregularities and insure the integrity of test results".

Table 1. Peer Review Guidelines for Test Security[2]

Critical Element	Examples of Evidence
2.5 The State has implemented and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results through: • Prevention of any assessment irregularities, including maintaining the security of test and administration procedures, incident reporting procedures, consequences for confirmed violations of test security, and requirements for annual training at the district and school levels for all individuals involved in test administration;	Evidence to support this critical element for the State's assessment system may include: • State Test Security Handbook; • Summary results or reports of internal or independent monitoring, audit, or evaluation of the State's test security policies, procedures and practices, if any.

Data Forensics

Data forensics analyses are used to identity statistical anomalies and testing irregularities that are indicative of potential test fraud. Consequently, it is recommended that these analyses be performed after every test administration in order to identify schools, teachers, test administrators, and students who might be associated with invalid scores and results.

Data forensic analyses should be supported by test delivery, scoring, and other systems that will capture and store appropriate test response data elements to make it possible to employ the detection statistics.

Data Forensics Statistics

Data Forensics Analysis includes the generation of various statistics designed to identify whether various test security breaches may have occurred. These analyses can be targeted to the specific threats identified by the State. Test security threats in K-12 assessment programs that can be detected by data forensics statistics include:

- Student test takers who share answers
- Teachers or other proctors who disclose the actual test questions, or allow proxy test taking

[2] Ibid., p. 30.

- Test content that may have been exposed prior to giving the test
- Test takers who may have been coached or received unauthorized assistance
- Test takers who may have communicated with each other during the exam
- Coaching of actual test content
- Disclosure of actual test content by an individual or on the Internet
- Inappropriate tampering of test materials, or inappropriate direction during testing

The RFP should indicate which test security threats are of concern; additional threats to those listed should also be considered. Respondents to the RFP should be asked to indicate and list the analyses that will be used to detect the threats that are listed in the RFP. The RFP should include unique aspects of the State assessment program, including the principles that govern forms construction and how the test will be administered. For example, when computers are used to administer tests, data in addition to selected responses may be collected and used. Examples of these additional data are to include when a student suspended the test to take a break, how much time was spent answering questions, whether students have changed answers to the questions and how often, the time of day when the test was administered, and/or the time when each question was answered. When available, these additional data elements can be processed and analyzed to detect potential test security violations. The RFP should detail these data elements and require responses from vendors as to whether these data will be collected and how they may be analyzed.

The RFP should indicate that the proposed data forensics analyses provide information about individual students, educators (through use of classroom and aggregated data), schools, and the exams. The RFP should ask how these analyses will be conducted to provide the requested information.

Use of Data Forensics Results

The data forensic analysis results should be sent to the State for review, should include recommendations for next steps, and should inform what appropriate responses may be invoked. A tight turn-around is often necessary to meet scoring and reporting deadlines following each administration of the tests. A procedure should be established for flagging identified scores following each administration with an appropriate status such as “indeterminate”, “hold”, “invalidation”, or other based upon the SEA’s policies and procedures.

Quality Assurance

Data forensics analyses can serve as quality assurance to ensure that policies and procedures are being implemented. Student flags may indicate the need for additional training for test administrators; school flags may suggest candidate sites for on-site monitoring in future administrations.

Peer Review Validity Evidence – Data Forensics Statistics

Critical Element 2.5 requires prevention, detection, remediation and investigation of testing irregularities.

Table 2: Data Forensics Evidence[3]

Critical Element	Examples of Evidence
2.5 The State has implemented and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results through: • Detection of test irregularities;	Evidence of procedures for detection of test irregularities includes documents such as: • Documentation of the information the State routinely collects and analyzes for test security purposes, such as description of post-administration data forensics analysis the State conducts (e.g., unusual score gains or losses, similarity analyses, erasure/answer change analyses, pattern analysis, person fit analyses, local outlier detection, unusual timing patterns).

Web and Media Monitoring

Given that exposure of content on social media is a major threat to the security of K-12 high stakes assessments, web and media monitoring services should help ensure that sensitive test information is not being communicated through the Internet and social media. Web and media monitoring can detect whether content is disclosed or at risk of disclosure through websites, peer-to-peer servers, social media, and other online channels. Web and media monitoring should search for English and potentially other language websites and searchable discussion forums for the disclosure of a State's protected test content and proxy testing solicitations. Reporting of findings should be delivered through weekly updates that detail the threats that have been identified and/or monitored. Each update should:

- Identify and classify each reported Internet risk (High, Medium, Low, for example)
- Track changes in risk status
- Report web traffic statistics for high-level risks
- Create a cloud-based archive of verified high-level risks, with URLs and other mutually agreed upon details of infringing content

Web and media monitoring services should be provided for a specified period around each test administration window. It is ideal to monitor one week prior to each administration, for each week during administration and for one week after each administration every contract year.

[3] Ibid.

Peer Review Validity Evidence – Web and Media Monitoring

Critical Element 2.5 requires prevention, detection, remediation, and investigation of testing irregularities.

Table 3: Testing Irregularity Evidence[4]

Critical Element	Example of Evidence
2.5 The State has implemented and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results through: • Detection of test irregularities;	Evidence of procedures for detection of test irregularities includes documents such as: • Documented incident reporting procedures such as a template and instructions for reporting test administration irregularities and security incidents for district, school and other personnel involved in test administration; • Documentation of the information the State routinely collects and analyzes for test security purposes, such as description of post-administration data forensics analysis the State conducts (e.g.; unusual score gains or losses, similarity analyses, erasure/answer change analyses, pattern analysis, person fit analyses, local outlier detection, unusual timing patterns); • Summary of test security incidents from the most recent year of test administration (e.g., types of incidents and frequency) and examples of how they were addressed, or other documentation that demonstrates that the State identifies, tracks, and resolves test irregularities.

Test Administration Monitoring

Test administration monitoring is a critical element of test security. On-site monitoring of test administrations can provide valuable information on the fidelity with which the test administration and security procedures are being applied. Monitoring procedures should include a plan for on-site monitoring of both paper/pencil and computer-based administrations, as well as the use of forms for documenting and certifying that applicable test administration and security procedures were followed by all school personnel. Technology that enhances the State's capability to monitor security procedures for assessment administration should also be included.

[4] Ibid.

A secure digital platform, configured to reflect the State's assessment policies, should be utilized to input, store, manage, analyze and report on data related to test security incidents. Levels of access to incident data should be provided based on staff requirements to protect confidential information. Data analytics and reporting should enable the State to easily comply with peer review guidance requiring States to be able to summarize incident data over time, including incident types and resolutions.

Peer Review Validity Evidence – Test Administration Monitoring

Critical Element 2.4 requires States to adequately monitor administrations of both regular and alternate assessments.

Table 4: Monitoring Evidence Requirements[5]

Critical Element	Example of Evidence
2.4 The State adequately monitors the administration of its State assessments to ensure that standardized test administration procedures are implemented with fidelity across districts and schools.	Evidence to support this critical element for the State's general assessments and AA-AAAS includes documentation such as: • Brief description of the State's approach to monitoring test administration (e.g., monitoring conducted by State staff, through regional centers, by districts with support from the State, or another approach); • Existing written documentation of the State's procedures for monitoring test administration across the State, including, for example, strategies for selection of districts and schools for monitoring, cycle for reaching schools and districts across the State, schedule monitoring, monitor's roles, and the responsibilities of key personnel; • Summary of the results of the State's monitoring of the most recent year of test administration in the State.

Investigation of testing Irregularities

Investigating testing irregularities is an important part of an overall test security quality assurance process. The purpose of conducting investigations is to enable the State to confidently determine the truth around a testing irregularity, including its complete scope and context.

Clear procedures need to be in place to guide the collection of credible evidence that documents the scope of an incident and any breach of secure materials, deviance from established policies, or individual

[5] Ibid.

misconduct. Any evidence gathered should be carefully preserved and reported to support further actions on the part of the local education agency (LEA) or SEA involved.

Test security investigations may include interviews with stakeholders of interest, including teachers, administrators, and potentially students. Interviews often shed the most light on the specific causes of irregularities and can help the State determine the circumstances surrounding a test irregularity. Specific protocols must be in place to guide any interviews during an investigation, particularly those that involve students.

When exploring the context of reported testing irregularities, States should have procedures to identify when experienced third-party professional investigators should be engaged to conduct a fair, thorough and unbiased investigation.

The results of irregularity investigations often provide important feedback about test security policies, training and practices both at the local and State levels. Investigations may uncover areas where existing policies may need to be refined or clarified to eliminate identified vulnerabilities. Investigations can also inform the scope and context of training and identify schools where future monitoring may be required, or pinpoint areas for more targeted data forensics analyses.

Peer Review Validity Evidence – Investigations of Testing Irregularities

Critical Element 2.5 specifically requires evidence of investigation of alleged or factual test irregularities.

Table 5: Investigations Evidence Requirement[6]

Critical Element	Examples of Evidence
2.5 The State has implemented and documented an appropriate set of policies and procedures to prevent test irregularities and ensure the integrity of test results through: • Investigation of alleged or factual irregularities.	Evidence of procedures for remediation of test irregularities includes documents such as: • A contingency plan that demonstrates that the State has a plan for how to respond to test security incidents and that addresses: • Different types of possible test security incidents (e.g., human, physical, electronic, or Internet-related), including those that require immediate action (e.g., items exposed online during the testing window); • Policies and procedures the State would use to address different types of test security incidents (e.g., continue vs. stop testing, retesting, replacing existing forms or items,

[6] Ibid, p. 31.

	excluding items from scoring, invalidating results); • Strategies for communicating with districts, schools and others, as appropriate, for addressing active test security investigations.

Appendix A. Additional Procurement Language

Security Audits and Handbooks

A Test Security Audit should include a review of all test security related documents as well as interviews with all staff responsible for test security at the SEA or organizational level. The Test Security Audit should cover the broad areas of test security defined by best practice using an industry accepted set of security standards and provide findings and recommendations for each area.

Security Audits and the development of Security Handbooks should follow industry best practices as defined in publications such as *CCSSO Operational Best Practices for Statewide Large-Scale Assessment Programs*, *The Handbook of Test Security*, and the *TILSA Test Security Guidebook*. Areas for a Security Audit should include:

1. Security planning
2. Test security roles and responsibilities
3. Budgeting and funding
4. Legal agreements
5. Test and item design
6. Test development and maintenance
7. Test publication
8. Test administration
9. Test scores and results
10. Physical security
11. Information security
12. Web and media monitoring
13. Security awareness and training
14. Managing security investigations

The Test Security Handbook should cover similar areas and contain references to existing policies as well as best practices for each heading:

1. Test security goals
2. Budgeting and finance
3. Legal precautions and agreements
4. Test distribution
5. Test administration
6. Test scoring and results
7. Physical security
8. Information security
9. Data privacy
10. Internet monitoring
11. Security breach action plan
12. Security awareness and training
13. Conducting security investigations

Data Forensics

State RFPs should spell out the requirements for data forensics to ensure that vendors can provide the level of detail and statistical analyses, reports, and report interpretation required to support state policies. Data requirements for handoffs between prime vendors and potential subcontractors should be identified and defined by the vendor(s).

The RFPs should also describe the specific types of analyses required given the anticipated security threats. Attention should also be paid to the modes of the administrations of interest (e.g. paper/pencil, computer based). RFPs should require specific analyses that would indicate non-independent test taking between pairs of students, collusion among groups of students or proctors and students, or prior exposure to test content. Data forensics analyses should also include analysis of unusual score gains or losses across classrooms or schools.

In addition, RFPs should also indicate whether score invalidations will be used and what reporting timelines are required to meet the State's reporting requirements. Vendors should also describe the process for working with the State to communicate data forensics results.

Additional detail on RFP language for Data Forensic Analysis is located in Appendix C and Appendix D.

Web and Media Monitoring

State RFPs should clearly identify the scope and duration of media monitoring with regard to the number of tests and the length of administration windows. Best practice includes monitoring one week before and one week after administration windows open and close. States should consider the relative risk of content disclosure and plan their media monitoring accordingly. For example, exposure of EOC content where tests are required for graduation may be a greater threat, with more significant consequences, than a grade 5, lower stakes, science assessment. RFPs should require vendors to describe their monitoring tools and processes, particularly the level of detail that will be provided to the State and procedures that will be used to remove any secure content that is detected.

Test Administration Monitoring

RFPs should clearly describe the scope and manner of monitoring State assessment administrations, as well as the systems used to capture and report monitoring data and observations. Examples of procurement language include:

- Standardized monitoring protocols and checklists should be developed in accordance with State testing regulations, policies and administration manuals to ensure careful, complete and fair monitoring of State assessment administrations across all schools.
- Two qualified monitors should visit each school selected by the State for monitoring. It is estimated that the duration of each school observation should range between 4 and 6 hours on site.
- Monitoring protocols should include the authentication of monitor credentials by school staff.

Vendors should provide comprehensive monitoring protocols that include requirements for staff assignments and training, scheduling, materials storage, document management and distribution, testing accommodations, testing rosters and other test administration documentation.

Monitors should identify and document any incomplete or missing documentation, proctoring issues and perceived violations of any policies and procedures. The monitoring protocols should include a reliable system for immediately notifying the State of any testing irregularities observed during monitoring, in accordance with criteria established by the State. Vendors should utilize a secure, Internet-based digital platform that allows monitors to enter monitoring data and observations in real time while on site at schools via electronic tablet devise in order to provide real time monitoring data to the State and promptly notify the State of any irregularities.

Monitors must be required to meet certain professional qualifications that will be established in consultation with the State and should receive training specific to the monitoring engagement. Monitor training should, at a minimum, include the goals of monitoring, relevant state laws and policies, examples or case studies, components of successful monitoring and instructions for recording and reporting data.

The secure monitoring platform should store all monitoring data collected by monitors, provide analytics on ever monitoring data point and allow for configurable reports to be generated based upon the data at the school, district or state level.

Investigation of Testing Irregularities

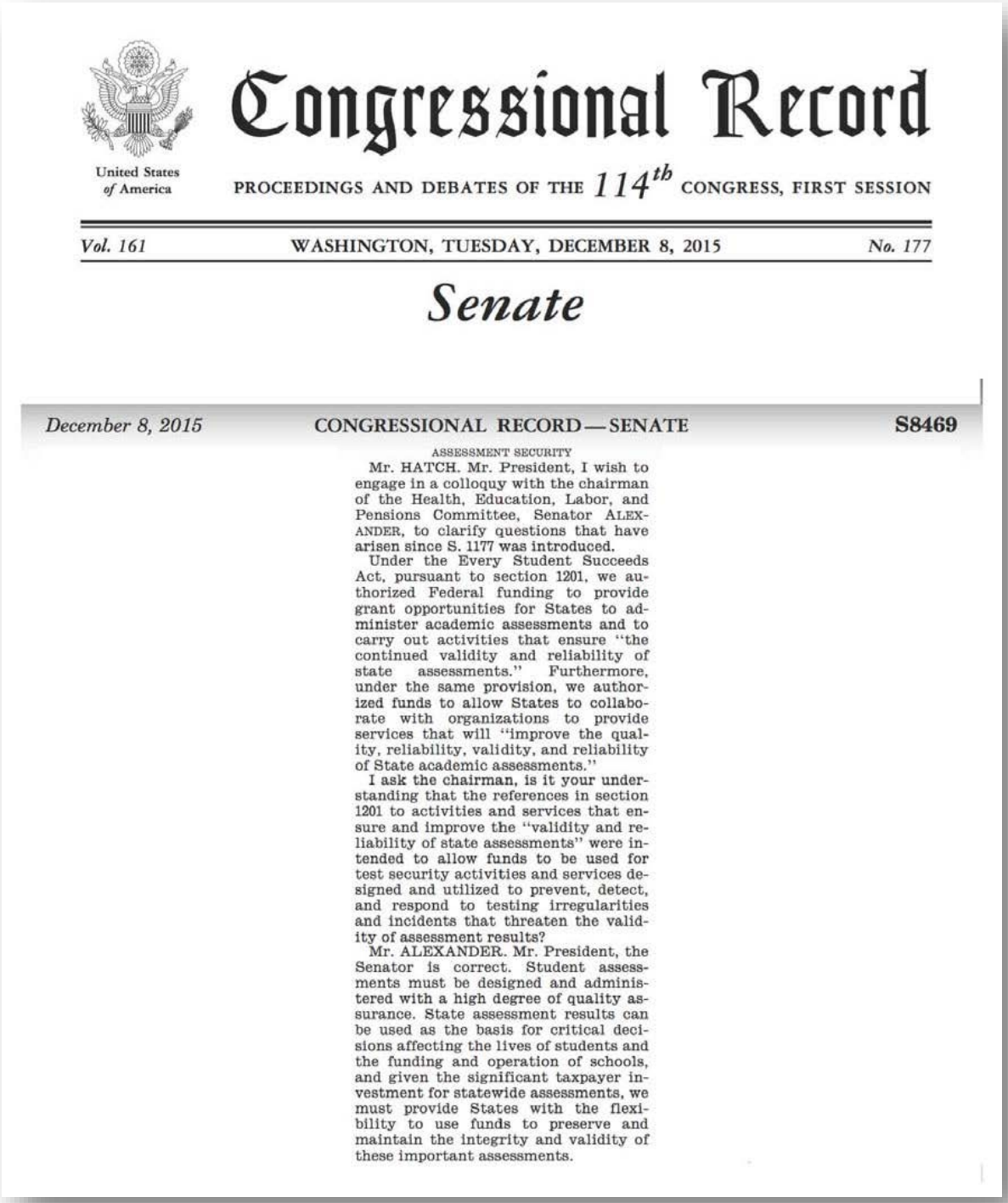
State RFPs should clearly identify the requirements of the systems used to report testing irregularities as well as the scope and manner of investigations of testing irregularities, and how the State will store and analyze all related evidence and report the findings of any investigations that may be conducted.

The vendor should provide a secure, Internet-based digital platform for schools and districts to promptly, completely and accurately report testing irregularities to the State in accordance with state laws and testing policies and store and analyze all data and evidence related to irregularities. The platform should allow for the classification of irregularities by severity and provide immediate notifications to the State when severe irregularities are reported. The platform should also provide automated analytics of all irregularity data and allows for configurable summary reports to be generated that show irregularities by type, classification and resolutions. Test irregularity investigations must be designed to obtain credible evidence that will enable the state to confidently determine the complete scope of the conduct at issue and identify all involved persons. Evidence gathered for this purpose should be collected, preserved and reported in a method that will enable the State to take any possible action it deems appropriate based upon an objective review of the evidence in accordance with state policies and law. Vendors should define a Security Incident Response Plan for dealing with each testing irregularity to be investigated. These plans should include the details of the incident, including involved personnel, a summary of the incident and available documentation and a plan for gathering additional facts and evidence.

The vendor must have an investigative procedure to identify and review all information reported in relation to the irregularity under investigation, as well as all data obtained during monitoring of assessment administrations at the school at issue and historical testing data from the school at issue, including, if available, data forensics analysis of test response data. The vendor must also have processes to identify school staff testing roles and room assignments, as well as related student identifying data in order to identify potential witnesses who may have knowledge of facts relevant to the irregularity under investigation.

Appendix B. Congressional Record Insert

Senate Colloquy on using Title I funds for test security



United States of America

Congressional Record

PROCEEDINGS AND DEBATES OF THE 114th CONGRESS, FIRST SESSION

Vol. 161 WASHINGTON, TUESDAY, DECEMBER 8, 2015 *No. 177*

Senate

December 8, 2015 CONGRESSIONAL RECORD — SENATE **S8469**

ASSESSMENT SECURITY

Mr. HATCH. Mr. President, I wish to engage in a colloquy with the chairman of the Health, Education, Labor, and Pensions Committee, Senator ALEXANDER, to clarify questions that have arisen since S. 1177 was introduced.

Under the Every Student Succeeds Act, pursuant to section 1201, we authorized Federal funding to provide grant opportunities for States to administer academic assessments and to carry out activities that ensure "the continued validity and reliability of state assessments." Furthermore, under the same provision, we authorized funds to allow States to collaborate with organizations to provide services that will "improve the quality, reliability, validity, and reliability of State academic assessments."

I ask the chairman, is it your understanding that the references in section 1201 to activities and services that ensure and improve the "validity and reliability of state assessments" were intended to allow funds to be used for test security activities and services designed and utilized to prevent, detect, and respond to testing irregularities and incidents that threaten the validity of assessment results?

Mr. ALEXANDER. Mr. President, the Senator is correct. Student assessments must be designed and administered with a high degree of quality assurance. State assessment results can be used as the basis for critical decisions affecting the lives of students and the funding and operation of schools, and given the significant taxpayer investment for statewide assessments, we must provide States with the flexibility to use funds to preserve and maintain the integrity and validity of these important assessments.

Appendix C. Additional RFP Language for Data Forensics Analysis

Introduction

Test fraud detection is essential for understanding, assessing, documenting, and handling test security threats for statewide assessment programs. The Department of Education (DoE) would like to use data forensic analysis of student test-response data from its statewide assessment program to identify patterns indicative of test fraud at the district, school, classroom, and individual student levels.

The results of these analyses will provide the DoE with critical information regarding where and when potentially suspect activity occurred, by whom, and its effects on the testing program. Using the results of these analyses will provide the DoE with ongoing information directly relevant to the fairness and validity of statewide assessment results.

Methods for Analyzing Data

In order to measure, monitor, and manage potential security issues with the test administration and to assess the validity of the test results, DoE requests that vendors conduct several statistical analyses to detect anomalies indicative of potential test fraud on the specified assessments. Statistics should be computed for online and paper and pencil tests, as appropriate:

- Similarity analysis
- Answer change/erasure analyses
- Timing and exam trend analysis
- Perfect test analysis
- Unusual score gains and drops

Similarity Analysis

The vendor should perform similarity analysis by comparing individual test instances with each other to identify improbable similarities between the responses on the tests. The analysis should identify pair of similar tests, clusters of similar tests, and schools or classrooms where anomalous levels of similar tests have occurred.

If vendor is proposing a Computer Adaptive Testing (CAT) solution, similarity analyses may be optional.

Answer Change/Erasure Analysis

The vendor should conduct an analysis of answer changes. Answer change analysis (a term typically used for computer-based tests) is also known as an erasure analysis when the analysis is conducted for paper-and-pencil tests. The analysis should measure unusualness of wrong-to-right (WTR) answer changes (i.e., replacing an incorrect answer with a correct answer) or score changes associated with answer changes (i.e., the difference between WTR and RTW erasures). Results should be tabulated for individual students, schools, and classrooms.

Timing and Exam Trend Analysis

The vendor should analyze test security changes and trends throughout testing windows. This analysis should indicate whether any exam content was inappropriately disclosed during the examination period. Minimally, a vendor's proposed timing and exam trend analysis should document changes during the test administration window of the following:

- Scores and pass rates,
- Administration volumes, and
- The number of detected anomalies.
- This analysis should be performed at the state level, district level, and school level.

Perfect Test Analysis

The vendor should analyze perfect tests, determining whether every response for a student's assessment is awarded the maximum possible score for the test question. This analysis should be conducted to assess whether the number of perfect tests was extreme for students within schools and classrooms.

Unusual Score Gains and Drops

The vendor should identify unusual score gains and drops for individual students and groups of students, where possible. At the student level, the analysis should compare each student's score with one or more scores from prior administrations. At the school and classroom level, the analysis should tabulate the extent of unusual score gains and drops that were observed within those groups.

Analysis of Groups

All the statistics described above should be computed for every student's test, if applicable. Some of the analyses are especially suited for the analysis of risk groups (i.e., groups where the probability of a security breach is higher than the norm). The following groups should be analyzed, where relevant data are available (additional groups, where applicable and agreed upon by vendor and the DoE, also may be analyzed):

- Schools
- Districts
- Classrooms/Teachers/Proctors
- Test Forms
- Individual Students

Data Forensics Reporting and Follow-up Support

The vendor should indicate the support that will be provided for actions that the DoE might pursue, which may include:

- Training for testing practitioners and proctors,
- Monitoring of future testing sessions,
- Auditing the physical security of testing environments,

- Academic probation,
- Educator reprimands and personnel file notations,
- Score invalidations, and
- Exam retakes.

On-Demand In-Depth Data Analysis

The vendor may be asked to provide further, more focused data forensics analysis, in areas where anomalies were detected or where patterns of repeated anomalies are detected such as at specific schools or within groups, with pricing to be determined on a case-by-case basis. The vendor should be prepared to help identify the nature of security violations that potentially occurred and to provide guidance for follow-up work to improve test and exam security and/or guidance for investigations into potential instances of test fraud.

Additional Requested Data Forensics Analyses

The vendor also is asked to provide a response to the following additional optional data forensics analyses, which may be performed at the discretion of the DoE.

- Synchronicity analysis
- Response time analysis
- Score difference analysis
- Person-fit statistics
- Excessive omitted responses, excessive multiple marks, and blank answer documents.
- Other statistics

Synchronicity Analysis

Synchronicity statistics analyze the time of day when students answered questions. The statistics detect when students answered many questions at or near the same time, indicating they may have been "paced" or "guided" during the test administration. This analysis utilizes response time stamps (i.e., the actual time when the question was answered) for the comparison, which can only be collected through computer-based tests, and thus the statistic is not applicable to paper and pencil tests. Only test instances for students who took the test during the same testing session should be compared.

Response Time Analysis

Response time statistics analyze the amount of time taken by students to answer questions. The statistics detect when students answer questions extremely quickly or in a manner that is inconsistent with typical test taking behavior. The analysis is only available if the response times are collected (i.e., through computer-based testing). Unusual response time patterns can indicate pre-knowledge of test content or unsanctioned aid given to students.

Score Difference Analyses

Score difference analysis, which is also known in the psychometric literature as differential person functioning, involves comparing each student's performance on subsets of items, such as the subset of scored items versus non-scored (i.e., field test) items or a subset of previously used items versus new items. Higher performance on scored items or previously exposed items could be indicative of pre-knowledge of those items. The vendor should provide a description of any scored difference analysis that may be performed. The vendor should be able to provide this information for individual test instances, classrooms, and schools.

Person-fit Statistics

Person-fit statistics document whether a student's test instance aligns with the testing model (e.g., item response or non-parametric model). These statistics identify discrepancies between the student's observed performance (i.e., selected responses) and expected performance (i.e., derived using the testing model). The vendor should provide a description of any person-fit statistics that may be performed.

Excessive Omitted Responses, Excessive Multiple Marks, and Blank Answer Documents

The vendor should perform an analysis to identify excessive omitted responses, excessive multiple marks, and/or blank answer documents.

Other Statistics

The vendor may choose to perform other statistical analyses not explicitly described here. The vendor should provide descriptions of the analyses to be performed.

Protection of Student Data

The vendor must describe security measures for securely transferring and maintaining test data.

Appendix D. Data Requirements for Data Forensic Analyses

Some specific data elements are required for the above security analyses. Others are not required, but requested. If possible, the requested elements should be provided. Required elements to be collected by the vendor and made available for data forensics analysis are presented herein in **bold** face font. Requested elements are presented in *italic* font.

Information about a Student

Student identifier – This identifier should be unique.

Other student information – These data can provide helpful information for assessing the test security risk and situation associated with a detected anomaly. Some of these data might be: name, age, gender, class standing, etc.

Information about the Exam

Exam name or identifier – This identifier is used for referencing and identifying the assessment.

Exam cut-score(s) – This value (or group of values) is used for placing a test score into a performance category (such as pass or fail). If an analysis of performance categories is important (it is usually critical for education exams which assign Adequate Yearly Progress), the assigned performance category must be provided with the test result data, or the exam cut-score(s) must be provided.

Form code – This value identifies the test form taken by each student. This code is required when multiple forms were administered to compare security risks between forms. When the exam has a form-specific variable section (i.e., group of items which are not scored but vary between forms) the form code will be needed to disambiguate the items.

Information about the Test Administration Location

School identifier – This documents the student's school and should also include the district identifier.

Classroom identifier – This identifier is used for grouping the data and determining whether security risks associated with the classroom were present. It also may be used in the similarity analysis for limiting the number of pairwise comparisons by comparing one test with another only when the two test instances were administered in the same classroom.

Information about the Testing Session

Date of testing session – These data are used for analyzing test results through time. They are especially useful for determining whether the security of the exam has deteriorated with increased exposure due to administering the same exam over multiple days or weeks.

Test score – The score on the exam is required. If it is not provided, it must be computed using an answer key. Generally, it is helpful to have both the test score and the answer key to validate the test score.

Test performance classification or category – This also may be known as the assignment of pass, fail, proficient, advanced, basic, or some other category which indicates an assessment of the student's competence. While not required, it is helpful to have this information.

Prior test scores – When these are provided, a score gain/drop analysis can be performed.

Testing session proctor and/or teacher – This identifier is used for grouping the data and determining whether security risks associated with the proctor and/or teacher were present. It also may be used in the similarity analysis for limiting the number of pairwise comparisons by comparing one test with another only when the two test instances were administered in the same classroom.

Testing session beginning and end time-of-day –The time of day the test session was started and when it was ended should be provided. Also, time of day of pauses and when the test session was resumed should be provided, if applicable. The time units should be expressed in the local time, rather than in the time of day where the server is located (e.g., servers located in data centers in different time zones). If possible, workstation times should be synchronized (e.g., eliminate differences due to workstation clocks being inaccurately set).

Testing session duration –The duration of the testing session can be used to diagnose potential test security problems such as when the tests are taken more quickly than a person could read the questions or when the test session extends beyond the allowed time limit.

Testing session time limit – This is the amount of time that was allowed for completing the test.

Testing session accommodations – These are usually provided as binary flags (e.g., 'yes' and 'no') and indicate variances in test administration procedures which have been allowed for a specific student. The most common accommodation is that of granting extra time for completing the test.

Test delivery method – It is often the case that all the tests are delivered in the same way (e.g., paper-and-pencil or by computer). However, when the tests are administered using two or more media, it can be very helpful to separate them by the test delivery method for verification purposes.

Other testing session information – Sometimes other important information such as the testing session registration code is associated with the test administration. If so, these data should be collected.

Exam Blueprints, Data Maps, and Other Data

The above data elements usually reside, or are provided, in vendor-specific formats. Often other program data are very useful or even required to properly process the exam data. Because these data elements are program specific, this section mentions some elements which could be very helpful in processing the data.

Response Data Mappings – Sometimes specific response data elements have special meanings. Examples of these are special codes to identify omitted responses or responses with multiple marks.

Answer keys and/or scoring rubrics – Answer keys and scoring rubrics are required to validate the scoring responses. Some of the analyses may require being able to differentiate between correct and incorrect responses.

Scored and non-scored flags – When some items are not scored, this information is required. Being able to verify and reproduce the test score is very helpful in validating that the data have been processed properly. Additionally, when non-scored items are present on the exam, other statistical analyses may be performed that may detect situations where the scored items were potentially disclosed.

Item format or type –The information is very helpful when verifying that the test result data have been processed properly.

Information about the Reponses to Questions

Item identifiers – These are unique codes that identify the items on the test.

Item responses – These are the student's list of responses to the test questions. These are required for any basic statistical analysis.

Response scores – These are usually provided as *0*'s and *1*'s, indicating whether the student's response was incorrect or correct.

Answer-changes – Answer changes can be recorded by computer for computer-based tests, or erasure marks on answer documents can be scanned during scoring of paper-and-pencil tests. When answer-changes are analyzed for tests delivered by paper-and-pencil they are known as "erasures." Answer-change data should minimally indicate whether the answer was changed from right-to-wrong (RTW), wrong-to-right (WTR), and wrong-to-wrong (WTW).

Response times or latencies – This is the number of seconds the student took to answer the question (usually only available when the test is given by computer).

Item review flags and counts – When tests are given by computer, it is possible to record the number of times an item was viewed. It also is possible to record sequencing information.

Item view information – related to item review *counts*, item view information includes the time of day the item was viewed (beginning time and end time), and the response provided (if any) during that view.

Item presentation order – Some security statistics may utilize the order in which the items were presented.

Response time stamps – This is the time of day the student responded to the item.

Item presented flag – When tests are given by computer, it may happen that the student never saw some questions. When this happens, it is important to disambiguate a true omitted response from a response for a question that was never seen

Table D-1: Checklist of Data Elements

Required elements are presented in **bold** face font. Requested elements are presented in *italic* font.

Included	
	Information about a Student
	Student identifier
	Other student information (such as name, age, and gender)
	Information about the Exam
	Exam name or identifier
	Exam cut-score(s)
	Form code
	Information about the Test Administration Location
	School identifier *Classroom identifier*
	Information about the Testing Session
	Date of testing session
	Test score (raw scores, percent-correct scores, scale scores, etc.)
	Test performance classification or category
	Prior test scores
	Testing session proctor and/or teacher
	Testing session beginning and end time of day (including pauses and when the test session was resumed)

	Testing session duration
	Testing session time limit
	Testing session accommodations
	Test delivery method (e.g., computer, paper-and-pencil, oral)
	Other testing session information (e.g., registration code or identifier)
Exam Blueprints, Data Maps, and Other Data	
	Response Data Mappings
	Answer keys and/or scoring rubrics
	Scored and non-scored flags
	Item format or type
Information about the Reponses to Questions	
	Item identifiers
	Item responses
	Response scores
	Answer-changes (such as right-to-wrong, wrong-to-right, and wrong-to-wrong)
	Response times or latencies
	Item review flags and counts
	Item view information (including beginning and end of view and response provided during the view)
	Item presentation order
	Response time stamps
	Item presented flag

References

Chief Council of State School Officers. (2013) *Operational Best Practices for Statewide Large-Scale Assessment Programs*. Washington, DC: Council of Chief State School Officers. 2013.

Olsen John and Fremer, John. (2013) *TILSA Test Security Guidebook: Preventing, Detecting and Investigating Test Security Irregularities*. Washington, DC: Council of Chief State School Officers. 2013.

U.S. Department of Education. "U.S. Department of Education Peer Review of State Assessment Non-Regulatory Guidance for States" (PDF). Ed.gov. U.S. Department of Education, 25 Sept. 2015. 2015.

Wollack, James A. and Fremer, John K., editors. (2011). *Handbook of Test Security*. New York: Routledge.

161 Cong. Rec. S177 (daily ed. Dec. 8, 2015) (colloquy statements of Sen. Hatch and Sen. Alexander.)