

Gregory Cizek
Wim van der Linden
Linda Cook

President
Vice-President
Past President

TESTING AND DATA INTEGRITY IN THE ADMINISTRATION OF STATEWIDE STUDENT ASSESSMENT PROGRAMS

October, 2012

Contributors to this document, listed alphabetically, include N. Scott Bishop (ACT, Inc.), Kristen Huff (USNY Regents Research Fund), Karen Mitchell (Association of Medical Colleges), Sherry Rose-Bond (Columbus City Schools, Columbus, OH), Paul Stemmer (Michigan Department of Education), E. Roger Trent (Consultant, Columbus, OH), and James Wollack (University of Wisconsin). We are grateful to all of the National Council on Measurement in Education members who took the time to comment on an earlier version of this document.

The NCME mission is to advance the science and practice of measurement in education.

Goals of the organization include:

1. Encourage scholarly development in educational measurement
 - a. Improve measurement instruments and procedures for their administration, scoring, interpretation, and use
 - b. Improve applications of measurement in assessment of individuals, groups, and evaluation of educational programs
2. Disseminate knowledge about educational measurement, including
 - a. Theory, techniques, and instrumentation for the measurement of educationally relevant human, institutional and social characteristics
 - b. Procedures appropriate to the interpretation and use of such techniques and instruments
 - c. Applications of educational measurement with individuals and groups
3. Increase NCME's influence within the educational measurement community to ensure sound and ethical measurement practices
4. Influence public policy and practice concerning educational measurement
5. Promote awareness of measurement in education as a field of study and work to encourage entry into the field and interdisciplinary collaboration
6. Provide members with a strong professional identity and intellectual home in educational measurement and enhance the value of membership in NCME
7. Increase the operating and financial capacity of the association to enhance its effectiveness and its national recognition

TESTING AND DATA INTEGRITY IN THE ADMINISTRATION OF STATEWIDE STUDENT ASSESSMENT PROGRAMS

Testing and data integrity on statewide assessments is defined as the establishment of a comprehensive set of policies and procedures for: a) the proper preparation of students, b) the management and administration of the test(s) that will lead to accurate and appropriate reporting of assessment results, and c) maintaining the security of assessment materials for future use. The policies must ensure that all students have had equal opportunities to show their knowledge, skills, and abilities and have been actively involved in demonstrating those opportunities through their engagement with the test. Educators, students, parents, school boards, legislators, researchers, and the public must have confidence that psychometrically-sound testing, scoring, and reporting will be handled ethically and in accordance with the best administrative practices to ensure that results accurately reflect *each student's own true educational knowledge, skills, and abilities*. For purposes of this document, we focus on the aspects of *test data integrity* that relate to maintaining test security and safeguarding against artificially inflated scores.

WHY TEST DATA INTEGRITY IS IMPORTANT

Federal¹, state, and local education decisions are based on results of statewide assessments. Assessment requires that results be: accurate, fair, useful, interpretable, and comparable. The technical merits of test scores must meet professional and industry standards with respect to fairness, reliability, and validity. Test data must be free from the effects of cheating and security breaches **and** represent the true achievement measures of students who are sufficiently and appropriately engaged in the test administration. Cheating, falsifying data, security breaches, and other actions of academic fraud compromise the standards of fairness, reliability, and validity by polluting data. When cheating occurs, the public loses confidence in the testing program and in the educational system which may have serious educational, fiscal, and political consequences. Policies and procedures must ensure that all students have appropriate, fair, and equal opportunities to show their knowledge, skills, and abilities. Students who need accommodations due to language differences or students with disabilities may require appropriate modifications to materials and administrative procedures to ensure fair access to the assessment of their skills.

WHO IS RESPONSIBLE FOR TEST DATA INTEGRITY?

Test data integrity is a shared responsibility among all educators, test professionals, and students². The ultimate leadership for ensuring data integrity belongs to State Educational Agencies (SEAs). However, Local Educational Agencies (LEAs) staff and students are critical partners in ensuring established test policies and procedures are properly implemented and followed. Assessment consortia, test publishers, and contractors also play a significant role. SEAs must have appropriate policies and legislation that address these issues, including descriptions of requirements, expectations, and consequences for assessment activities. LEA policies and procedures must address how data integrity is ensured within each district and school.

¹ The U.S. Education Department (ED) sets policy for score use in federal programs. ED can help ensure that legislation and rules governing test security are established by states and that there is appropriate consistency across entities. ED might also consider establishment of a repository for policies, rules and best practices that will help SEAs and LEAs ensure data integrity.

² For an example of ethical standards, see NCME's *Code of Professional Responsibilities in Educational Measurement* at the following link:

<http://www.ncme.org/resources/code.cfm>

RECOMMENDED PRACTICES

1. Entities should develop a comprehensive data integrity policy to ensure the fairness, reliability, validity, and comparability of results when tests and results are used as intended³. The policy should define assessment integrity (and why it is important) and set forth standardized practices that are practical within typical school environments, resources, and operations. It should define proper and prohibited conduct and include how to prevent irregularities. It should establish required security guidelines for protecting test materials from security breaches (where students who have not taken the test would get access to questions) and preserve questions for future use. School personnel should provide input during policy development and be given ample lead time for implementation before any new policy becomes effective.

Implementation plans should be tailored to the purpose of testing, how test scores will be used, and the format of test administration⁴. The policy should describe specific, required security measures, testing procedures, and testing conditions. Clear and consistent written procedures should describe preventive actions, appropriate and inappropriate actions, communication plans, and remediation steps.

The following points should be covered in the policy⁵: staff training and professional development, maintaining security of materials and other prevention activities, appropriate and inappropriate test preparation and test administration activities, data collection and forensic analyses, incident reporting, investigation, enforcement, and consequences. Further, the policy should document the staff authorized to respond to questions about the policy and outline the roles and responsibilities of individuals if a test security breach arises. The policy should also have a communication and remediation response plan in place (if, when, how, who) for contacting impacted parties, correcting the problem and communicating with media in a transparent manner.

2. Assessment Consortia, State Educational Agencies (SEA) and Local Educational Agencies (LEA), including school districts, and building administrative staff, should develop and implement appropriate training in proper administrative procedures and methods to prevent test irregularities. Training should provide an overview of ethical and proper administration procedures and stress the importance of academic and assessment integrity as a means of avoiding serious negative consequences for the testing program and its potential damage to the educational reputation of students and schools. Staff and students should understand and support monitoring efforts to report and detect breaches of security, cheating, and other improper behavior.

Training materials should address the difference in secure and non-secure testing materials (e.g., released materials, practice materials, etc.) and provide clear examples of what behavior is unacceptable during and after testing⁶.

Finally, training should ensure that staff and students are aware of the consequences if they are found to have engaged in conduct that threatens the integrity of test administration and results. Procedures to be followed in the event of a staff member or student being accused of misconduct should be articulated and reviewed in training. The procedures should address the appropriate understanding and

³ SEAs, LEAs, and schools must disseminate this information to all staff who participate in testing. Roles and responsibilities should be aligned (i.e., the SEA's plan will drive the LEA responsibilities, and in turn, the LEA's plan will drive the school's).

⁴ Threats for an end-of-course computerized test are different than those for a paper-and-pencil test used for accountability. Testing practices change (e.g., pencil and paper tests may become computerized), so data integrity plans will need to be updated accordingly.

⁵ More information and resources that may be helpful for developing these policies are provided in the Appendices. Consider utilizing technical advisors (e.g., SEA technical advisory groups) to vet the plans. Peer review processes might also be considered.

⁶ See Appendix A for some examples.

compliance with nondisclosure and confidentiality agreements, as well as participation forms for verifying that staff have participated in training. The expectation of compliance with administration standards should also be made clear to students. Older students might be asked to sign assessment conduct and responsibility statements as well.

3. Entities should engage in proactive prevention to minimize threats to data integrity. One source of cheating by staff is lack of understanding about what are acceptable and unacceptable behaviors and the important reasons behind the need for accurate test results. Efforts should be taken to eliminate opportunities for test takers to attain scores by fraudulent means, or opportunities for school staff or other stakeholders to tamper (violate instructions for appropriate administration or accommodations) with computer-based testing systems, paper-based test booklets, answer documents and other secure materials and information. Monitoring programs where operational assessments are observed by SEA agents also helps ensure assessment integrity⁷. Results of monitoring should be used for prevention and training (feedback to the school) as well as to identify potential irregularities.

Students should be told about the importance of the assessments and why it is important that the scores reflect their true abilities.

4. Entities should ensure that all test administrations follow standardized procedures as appropriate to the student (e.g., some students may require accommodations) and in accordance with the Standards for Educational and Psychological Testing (1999) or any of its subsequent revisions. Any and all guidelines regarding materials prohibited in testing areas should be followed⁸.

5. A clear and fair monitoring and investigation process to identify irregularities must be established by the SEA and a local version by each LEA. Entities should ensure all evidence of irregularities that are collected are comprehensive and facilitate subsequent analyses. This should include a detailed record of test administrators, support staff (proctors), and teachers' names. The requirements for data files used for integrity analysis will likely evolve as analytic techniques evolve⁹. In documenting irregularities, collection of physical evidence (e.g., cheat sheets), photographic evidence (e.g., notes written on arm, desk, etc.), examinee handwriting in test booklets or scratch paper, and other specific observational notes can play an important role during follow-up investigations.

For computer-based testing, Internet activities should be monitored and logged (sites visited, screenshots taken, etc.) for all persons who access school and district servers and the activities of all users of school/district computers. Computers should be checked for prohibited software and malicious programs.

6. Entities (e.g., SEAs or their designees) should conduct comprehensive integrity analyses at multiple levels (e.g., district, school, classroom, and/or students) for all large-scale programs where consequences for students and/or school personnel are present. State results typically provide the best comparison for evaluating schools and districts. Such analyses and reports should be reviewed by the SEA's technical advisory

⁷ Other preventative suggestions are provided in Appendix B.

⁸ See Appendix C for some examples of materials students have used to gain unfair advantage over others.

⁹ See Appendix D.

panel. The analyses should include multiple methods and follow best practices to ensure the highest likelihood of detecting misconduct, while using appropriate statistical controls to minimize false detections. Results should only identify students, classes, schools, and districts where there is strong evidence that further investigation for possible improprieties is warranted. Investigations and subsequent actions should focus on appropriate remediation and future prevention of any irregularities discovered.

7. In the interest of protecting the privacy of both those being investigated for potential cheating and those contributing information to the investigation, entities should ensure that reports of suspected cheating, security breaches, as well as other suspicious activities are developed following clear and transparent guidelines, and in accordance with the Freedom of Information Act, Family Educational Rights and Privacy Act and other applicable laws or professional guidelines. Individuals who report suspected violations must be protected from retribution. Multiple reporting avenues (e.g., 800 numbers, e-mail, web forms, etc.) should be provided. Clear methods, procedures, data analysis and findings and reports should be thoroughly documented. A secure database collections system for capturing reported incidents should be created and maintained. Appropriate sections of the system should be made accessible to all LEAs.

8. Entities should ensure the appropriate investigation of any reported incidents and irregularities that are flagged during forensics analysis. Qualified and trained staff responsible for investigating violations should be identified in advance. The SEA should develop policies for when and how to turn investigations over to a third party so as to avoid potential conflicts of interest. Investigations should occur in a timely fashion and written reports should be given to the SEA along with remediation plans for any problem areas.

9. SEAs and LEAs must develop plans to enforce breaches of assessment integrity and to handle the consequences in a fair and appropriate manner and most importantly, to ensure that the offense does not happen in the future. Sanctions or remediation must be proportional relative to the offense and equivalent to other policies. All parties should create and maintain due process and appeal procedures for suspect students and staff. The accused should be informed of the allegations or complaints and the circumstances behind them (statistical detection, reported violation, etc.).

10. As testing technology evolves, security needs and how we define test and data integrity must keep pace. Policies and procedures should be reviewed to ensure compliance with the principles of assessment integrity. Computer-based testing will present different challenges based on the hardware (mobile vs. desktop configurations), the software, and Internet configurations (network security, social media, etc.). A few examples include greater accessibility to biometric identification procedures, built-in universal design, handwriting analysis, time-stamping items and events, video/audio monitoring systems, and improved real-time and post-hoc statistical anomaly detection techniques.

APPENDIX A: SOME THREATS TO TEST INTEGRITY

The following is a non-exhaustive list of examples which have the potential to artificially inflate test scores.

Before Testing

- Using actual or live test items in continuous drilling instead of focusing on assessing the underlying learning standards
- Using secure/unreleased items to train students that violates the administration manual guidelines
- Previewing the test before administration
- Excluding selected students from the administration (e.g., not allowing lower-achieving students to sit for an exam in order to raise group averages)
- Using unauthorized test preparation materials
- Failure to store secure test materials
- Improper or ineffective test administration training practices (failure to train staff, failure to devise effective practices)

During Testing

- Students copying answers from other students
- Students providing assistance to or accepting assistance from other students
- Students or teachers using prearranged signals (e.g., tapping, signing, voice inflection, facial expression) to provide correct answers to students
- Failing to follow prescribed test administration procedures leading to administration irregularities, e.g., incomplete student responses, or providing too much information so as to assist the students in correctly answering questions
- Inappropriate proctoring by coaching or signaling students (e.g., hints, rephrasing questions, voice or facial inflection), pointing out errors, or otherwise identifying correct answers during the exam
- Displaying improper information in student assessment rooms
 - Putting up posters or other materials that provide test answers
 - Failing to cover existing information boards, posters
- Giving unauthorized students extended time, prohibited materials, or other non-standard conditions.
- Allowing unauthorized people in the testing area (e.g., media, other students, teachers, or parents)
- Inappropriate or over-accommodated student accommodation practices

After Testing

- Altering student answer documents, changing answers, or filling in omitted items
- Falsifying identification or demographic information for students
- Exposing or releasing items that will appear on future test forms
- Divulging details about test items to others who have yet to test (note: school staff should explicitly instruct students not to do this)
- For performance-based assessments, allowing local scoring that may favor responses of local students or staff scoring their own students
- During reporting, inaccurately summarizing or interpreting test results to the students' advantage
- Not returning all secure testing material
- Photocopying, reproducing, disclosing, or disseminating testing materials in any way
- Failing to submit answer sheets for students expected to do poorly
- Any other action resulting in data that misrepresents the achievement levels of students within classes, schools, districts, and states

APPENDIX B: SOME PREVENTIVE ACTIONS

The following is a non-exhaustive list of examples.

PAPER-AND-PENCIL ADMINISTRATION

Security of Materials

- Keep sensitive test materials (live test items and booklets, computer screens, or computer testing access, etc.) secure and accounted for at all times (before, during, and after testing)
 - Have a dedicated, secure place to store materials that prevents non-authorized access to test material
 - Determine which staffers have legitimate access to the storage area
 - If the storage area cannot be completely sequestered, track all staff who enter/exit the area
- Determine which staffers are responsible for maintaining the chain of custody over test materials (this applies to all administrative staff who handle test and proctoring materials)
- Pre-seal booklets (sometimes cost-prohibitive) or provide self-seals for students' test documents

Distribution and Collection of Materials

- Schedule the times that materials will be distributed and collected
- Specify and document check in/check out procedures for materials
- Promulgate a list of detailed procedures for reporting missing and damaged test materials

Test Administration

- Use seating charts and assign seating, as appropriate
- Require appropriate identification or recognition of each student as appropriate
- Seat students an appropriate distance apart
- Restrict or prohibit (as your administration manual requires) mobile cameras, cell phones, and other similar devices
- Use only trained test proctors and provide proper supervision (use proctor guidelines)
- Establish qualifications requirements (i.e., education and credential) for proctors and test administrators
- Have rooms proctored during the entire administration
- Document proctor names and locations of the assessment
- Independently monitor test administrations on a random basis
- Test all eligible students
- Do not allow teachers to test their own students unless necessary or allowed for by required accommodations
- Maintain established security procedures throughout make-up testing and special accommodations
- Establish common scheduling time and calendar for testing
- Have materials returned immediately after testing
- Test all examinees in a narrow testing window, scheduling primary subject matter tests on the same day and at the same time to reduce possible collusion and mitigate damages from a security breach
- Clearly identify prohibited behavior and items as well as rules for handling irregularities

COMPUTER-BASED ADMINISTRATION

Security of Materials

- Keep screens out of view of each student or others (position monitors, cardboard screens, and carrels strategically)
- Establish a building testing schedule so all students are tested in the same subject before beginning the next subject
- Time-stamp all student and staff access
- Specify disallowed access times (i.e., weekends, holidays, after hours, etc.)
- Ensure that students are locked out from accessing unauthorized computer applications, including the use of the Internet, during assessment
- Lock-out access to the test after testing windows are completed
- Prohibit students from accessing memory storage or Wi-Fi on mobile devices

APPENDIX C: SOME MATERIALS ALLOWED AND PROHIBITED DURING EXAMS

The following is a non-exhaustive list of examples. Always consult your specific administration manual.

Items Frequently Allowed in Testing Areas

- Admission ticket
- School-issued ID
- Government-issued photo ID
- Number two pencils (wooden)
- Quality erasers
- Highlighters, other approved writing implements
- Silent or beeping timers
- Foam ear plugs or other noise-blocking devices
- Transparent containers (e.g., “Ziploc bags”)
- Approved calculators
- Water bottles, as approved
- Dictionaries, as approved

Items Frequently Prohibited in Testing Areas

- All electronic devices used for communication or data storage (e.g., cell phones, book readers, tablets, pagers, cameras, non-approved calculators, music players, voice recorders, etc.)
- Study, review, or other information resource materials (dictionaries, thesauruses, encyclopedias, spelling and grammar checkers)
- Correction fluid, correction pens
- Large rubber bands, large pencil erasers
- Boxes, pencil cases, eyeglass cases, or other opaque containers
- Briefcases, backpacks, purses
- Clothing that could be disruptive or present a potential test or student security threat (e.g., hats, scarves, hoodies, loose or bulky clothing)
- Earphones, headphones, ear buds unless as a required accommodation or computer administration requirement
- Mechanical pencils or ink pens (except for notes for computer-based testing or other exceptions)
- Smoking materials, food, beverages (Note: case-by-case exceptions for medical reasons can be made)

APPENDIX D: DATA COLLECTION AND ANALYSIS

Forensics should be considered carefully and determined as appropriate for each test by Technical Advisory consults and/or committees. Analysis should be technically sound and carefully targeted to avoid false positives, while simultaneously maximizing true positives. Suggestions of what to collect and look for include the following:

Suggestions for Data Collection

- Capture both teacher and proctor names (e.g., on classroom ‘header’ sheets) and include this info in data files for potential use in forensic analyses
- Expand the contents of the data file(s) used for integrity analysis by including:
 - actual student scan/scored vectors (e.g., A, B, C, D for ‘wrongs,’ 1, 2, 3, 4 for ‘rights’)
 - ability information (raw and/or scaled scores)
 - pre-erasure answer strings
 - post-erasure answer strings
 - string of erasure types (wrong-to-right, wrong-to-wrong, right-to-wrong, no erasure)
 - darkness gradient for post-erasure item responses
 - pixel coverage of post-erasure item responses

Suggestions for Forensics Analysis

- Suspicious changes in test scores in adjoining test years
- Suspicious changes in student demographics across years
- Suspicious erasures
 - high erasure rates and, in particular, high wrong-to-right erasures
 - erasures with different darkness and pixel coverage than non-erased responses.
 - contrast erasure rates for pilot versus operational items
 - consistency of erasures (i.e., erasures on the same set of items) for students within classrooms, schools, and districts versus the state
- Speed of responding on computer-based tests
- Similar answer patterns between pairs or groups of students
- Similar items being flagged as erased between groups of students
- Similar responses to open-ended items
- Inconsistent item responses pattern—response aberrations, in particular for pre- and post-erasure responses
- Outliers in scatter plots of subject area scores (e.g., what classes had mathematics scores that were outliers based on reading score performance)
- Prior test administration common items (e.g., one year back) vs. common items from several years prior (multiple years back) as well as comparison between operational and pilot sections may help identify students who had pre-knowledge of questions
- Comparisons between summative assessments and earlier formative/interim assessments, third party assessments, such as NAEP, or other academic efforts (GPA, class rank, coursework)

APPENDIX E: RESOURCES

Policies and procedures must be based on best practices in testing. Some of these documents are showing their age and are in various stages of revision. Among the documents to be considered in establishing the definitions and descriptions of best practices are:

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students*. Washington, DC: NCME.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

National Council on Measurement in Education (1995). *Code of Professional Responsibilities in Educational Measurement*. Washington, DC: Author.

Joint Committee on Testing Practices (2004). *Code of Fair Testing Practices in Education*. Washington, DC: American Psychological Association.

Council of Chief State School Officers and the Association of Test Publishers (2010). *Operational Best Practices for Statewide Large-Scale Assessment Programs*. Washington, DC.