



THE STATE EDUCATION DEPARTMENT / THE UNIVERSITY OF THE STATE OF NEW YORK / ALBANY,
NY 12234

TO: P-12 Education Committee

FROM: Valerie Grey
Ken Slentz

SUBJECT: Proposals to improve integrity of the State's tests

DATE: October 13, 2011

AUTHORIZATION(S):

SUMMARY

Issue for Action

At the September meeting, the Board charged the SED testing workgroup to provide additional information on proposals to improve the integrity of the State's tests.

Reason(s) for Consideration

Review and approval.

Proposed Handling

This topic will come before the P-12 Education Committee at its October meeting, where options will be discussed and action taken.

Background Information

At the September 12, 2011 meeting of the P-12 Education Committee, the Board discussed how to best ensure the integrity of the 3-8 assessments and Regents exams in order to continue to: (1) accurately measure student performance; and (2) develop our testing program into the sophisticated and rigorous system necessary to support meaningful education reform. The Board approved pursuing an independent review of the Department's procedures related to incident reporting and follow-up on allegations of testing impropriety. Approval was given to further develop, for consideration at the October 2011 Regents meeting, specific proposals on proctoring, scanning and scoring of the State assessments.

In order to gain a better perspective on the current practices in scanning, scoring and ensuring test integrity, the workgroup obtained information from various states throughout the country on their practices. The majority of the states reported using outside vendors to manage scanning and scoring and distribution and collection of testing materials. Additionally, many states are shifting their efforts toward computer-based assessments. Some states currently administer at least some accountability assessments on computers and use statistical measures to detect testing irregularities.

New York is planning to use computer-based testing beginning in the 2014-15 school year. Currently, the Partnership for the Assessment of Readiness for College and Careers (PARCC) consortium¹ is evaluating the appropriate tools, technology, and processes that will be required for full implementation. PARCC is also working to make a technology readiness tool available to member states in early 2012. We expect that information will be available this spring regarding the processes and tools that PARCC will procure for test administration and scoring. This information will be shared with the Board as the level of effort and resources required to adopt PARCC is evaluated.

Scanning/Scoring Multiple Choice Answer and Erasure Analysis

As discussed at the September Board meeting, a statewide centralized scanning and scoring system has a number of advantages. Centralized scanning of multiple choice item responses can provide rapid return of test scores and employ consistent readily applied methods for detection of testing irregularities. It is estimated that creating a statewide centralized scanning system that includes erasure (for pencil responses) and wrong-answer analysis (for ink responses) would cost the State approximately \$10 - 12 million annually.

New York is unique in its reliance on local scanning and scoring, and significant investments have been made at the local level to develop infrastructure. Currently, scanning occurs at the regional level through Regional Information Centers (RICs) and Big 5 city school district scan centers. Last year, in response to an Office of the State Comptroller (OSC) audit, the Board approved new scanning requirements for Regents exams. These new rules were phased in and will be fully implemented in June 2012. SED is continuing to explore the capacity and timeliness of scoring and erasure and ink analysis at the local level, particularly for the Regents exams, and will provide a report back to the Board later this year.

In the past, the Department contracted for limited erasure analysis to help ensure validity, particularly when rolling out new Regents examinations. In 2005, the Department contracted with Pearson Inc. for the post-administration scanning and collection of data from new exams as part of its validity processing. Subsequent

¹ In January, 2010, the Regents endorsed the participation of New York State in the 24-state Partnership for the Assessment of Readiness for College and Careers (PARCC). PARCC is a consortium of states that worked together on a joint proposal to USDE to seek Race to the Top funding for the development of a K-12 assessment system aligned to the Common Core State Standards in English language arts and mathematics for grades 3 - 11. PARCC was awarded a total of \$185 million in September 2010.

amendments to the contract included, among other provisions, erasure analysis for two administrations of four new Regents exams. Overall, since 2008, erasure analysis in some form was completed on eight administrations of four new Regents exams out of 229 Regents exams administered during this period.

While the purpose of the erasure analysis was not specifically to identify testing improprieties, Department staff used the data to identify 64 schools with erasure patterns that warranted additional scrutiny. After further review, seven of those 64 schools had findings that called for full investigations, and the Department referred those cases to the appropriate investigating authorities.

It is recommended that the Department perform erasure analysis (for pencil responses) and wrong-answer analysis (for ink responses) for 10% of Spring 2012 3-8 assessments and Regents exams. This is estimated to cost approximately \$1 million and funding would be requested in the SFY2012-13 State Budget. This additional funding would ensure erasure and wrong-answer analysis for roughly 500,000 answer sheets, an increase of nearly three-fold from the erasure analysis done for new test validity last year.

Open Response Scoring

Beginning with the 2010-11 school year, schools were no longer permitted to rescore any open-ended questions on Regents exams, a longstanding practice that may have contributed to a statistically improbable grouping of scores around the key passing marks of 55 and 65. Beginning in the 2012-13 school year, it is recommended that all teachers be prohibited from scoring their own student's exams. This universal prohibition will likely necessitate a change in the testing calendar for next year's Regents exams.

Districts must develop the capacity to implement alternative scoring processes, such as, external, regional, or distributed technology for the scoring of open responses on State assessments. This will be challenging. However, there are examples of how it can be done successfully throughout the state. In Ulster BOCES, seven of eight districts participate in regional scoring for 3-8 assessments. In eastern Suffolk BOCES, regional scoring is done for Regents exams. New York City is currently exploring a distributed scoring platform (an electronic means of sharing open response answer documents for independent scoring by non-vested individuals). Other cities, like Yonkers, are currently using a version of distributed scoring. SED will work with districts that choose to build a distributed scoring platform, to the extent possible, to adjust certain testing parameters to ease cost/operations at the local level.

It is recommended that the Department amend existing contracts to provide additional inter-rater reliability studies for 3-8 assessments and Regents exams. Inter-rater reliability studies are conducted after tests have been administered to measure the consistency with which two or more scorers rate the work or performance of test takers on open response items. These studies provide important evidence regarding the quality of test scoring and the appropriate application of scoring rubrics. This is estimated to cost approximately \$0.7 million and funding would be requested in the SFY2012-13 State Budget to enhance our current capacity for this analysis.

Computer Based Testing Pilot

A strategic shift toward computer-based testing will help resolve many of the test integrity issues described above, while also allowing for just-in-time assessment results and the use of innovative assessment items that are aligned with career and college readiness. While this shift will require a major investment in the state's computer and broadband network infrastructure, this initiative will have long-term advantages, not only for the administration of assessments, but also by supporting innovative uses of educational technology to support instructional improvement. New York lags behind many states in its capacity to support computer-based testing. It is recommended that funding of \$200,000 be requested in the 2012-13 State Budget to pilot a digitized test to learn more about the benefits of this testing model, the possible cost efficiencies, and the challenges that the State and districts will face in implementation. To aid in the identification of these challenges, the PARCC consortium will provide a technology needs analysis tool to all PARCC states early in 2012.

Data Forensics

Using better detection methods, SED can both investigate potential abuses of the system and deter cheating in the future. In their book *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, Brian A. Jacob from Harvard University and Steven D. Levitt from the University of Chicago conducted a comprehensive statistical analysis of testing in Chicago. This type of data analysis, as well as others, could be undertaken by partnering with researchers and scholars to improve our ability to identify testing irregularities. It is recommended that funding of \$200,000 be requested in the 2012-13 State Budget to support these efforts. The experts serving on the Department's Technical Advisory Group will be consulted to assist with this work.

Independent Reviewer

The Department is in discussions with several possible independent reviewers to look at the Department's procedures related to incident reporting and follow up of allegations of testing impropriety at schools. Selection of the reviewer will be made in the next couple of weeks. It is possible that the reviewer will make recommendations that could result in 2012-13 budget requests in addition to those outlined in this document.

Retention of Answer Sheets

Schools are currently required to retain answer sheets for one year. Requiring a longer retention period to ensure investigative capacity was discussed at last month's meeting of the Board. This issue is particularly relevant to the work expected to be done by the Independent Reviewer, and we will request that this individual make a recommendation for consideration by the Board.

Proctoring Procedures

At the September meeting, the Board asked the workgroup to further explore a potential prohibition on teachers from proctoring the exams of their own students or other exams in their certification area. After further research, the Department found that most states we contacted do allow teachers to proctor their own students' tests. This, combined with the possible impact on younger students, leads us to recommend no change to current proctoring policies.

Local Monitoring Best Practices and Requirements

Some districts more aggressively monitor testing than others by providing on-site inspection of test administration. This is an area where it makes sense to identify best practices and develop state requirements for monitoring protocols. It is recommended that the Board direct Department staff to quickly form a small representative workgroup to get feedback and make specific recommendations in November, leaving time for districts to become familiar and meet any new requirements starting with the January 2012 administration of Regents exams.

Recommendations Summary

The testing workgroup will continue its work and the following actions are before the Board for approval:

For school year 2011-12:

- Request funding of \$1.0 million in the 2012-13 State Budget to perform scanning/erasure analysis on 10% of 3-8 assessments and Regents exams post administration for multiple choice;
- Request funding of \$0.7 million in the 2012-13 State Budget to support additional inter-rater reliability analysis of open response scoring on 3-8 assessments and Regents exams;
- Request funding of \$0.2 million to plan and begin a computer based testing (CBT) pilot.
- Request funding of \$0.2 million to support additional data forensics, in consultation with the Technical Advisory Group, and in partnership with academic partners.
- Create a small working group to make recommendations for local monitoring protocol requirements;
- Request Independent Reviewer make recommendations on appropriate retention time of answer sheets and assessment booklets.

For school year 2012-13:

- Prohibit teachers from scoring their own student's exams. Districts must develop the capacity to implement alternative scoring processes, such as, external, regional, or distributed technology for the scoring of open responses on State assessments.