



State Education Department

Oversight of Scoring Practices on Regents Examinations

Report 2008-S-151



Thomas P. DiNapoli

Table of Contents

| | Page |
|--|------|
| Authority Letter | 5 |
| Executive Summary | 7 |
| Introduction | 9 |
| Background | 9 |
| Audit Scope and Methodology | 10 |
| Authority | 10 |
| Contributors to the Report | 11 |
| Audit Findings and Recommendations | 13 |
| SED Oversight | 13 |
| Recommendations | 17 |
| Training in Examination Scoring | 18 |
| Recommendations | 19 |
| Agency Comments | 21 |
| State Comptroller's Comments | 27 |

State of New York Office of the State Comptroller

Division of State Government Accountability

November 19, 2009

Mr. David Steiner
Commissioner
State Education Department
State Education Building - Room 111
89 Washington Avenue
Albany, NY 12234

Dear Commissioner Steiner:

The Office of the State Comptroller is committed to helping State agencies, public authorities and local government agencies manage government resources efficiently and effectively and, by so doing, providing accountability for tax dollars spent to support government operations. The Comptroller oversees the fiscal affairs of State agencies, public authorities and local government agencies, as well as their compliance with relevant statutes and their observance of good business practices. This fiscal oversight is accomplished, in part, through our audits, which identify opportunities for improving operations. Audits can also identify strategies for reducing costs and strengthening controls that are intended to safeguard assets.

Following is a report of our audit of the State Education Department's Oversight of Scoring Practices on Regents Examinations. This audit was performed pursuant to the State Comptroller's authority under Article V, Section 1 of the State Constitution and Article II, Section 8 of the State Finance Law.

This audit's results and recommendations are resources for you to use in effectively managing your operations and in meeting the expectations of taxpayers. If you have any questions about this report, please feel free to contact us.

Respectfully submitted,

*Office of the State Comptroller
Division of State Government Accountability*



State of New York Office of the State Comptroller

EXECUTIVE SUMMARY

Audit Objective

Our objective was to determine whether State Education Department (SED) oversight of local school districts provides adequate assurance that local districts accurately score Regents exams.

Audit Results

Regents exams are statewide high school tests in particular subject areas. The exams are to provide reliable measures of academic performance for each student and for entire schools. SED develops the exam questions and answers and distributes these to local school districts under secure procedures. Local school districts administer and grade the exams using SED instructions, answer keys for multiple choice questions and scoring guidelines for questions involving judgment. Local school districts provide exam results to SED for oversight purposes including analyzing trends and reporting statewide academic performance. SED oversight also includes selectively obtaining the scored examination for review of the accuracy of scoring throughout the State.

SED reviews of the scoring of selective Regents exams have identified significant inaccuracies by local school districts. These inaccuracies have tended to inflate the academic performance of students and schools. While SED has detected this problem, its oversight has not been adequate to ensure that local school districts will correct the problem so that future exams are more accurately scored.

For example, a team of SED experts rescored certain June 2005 Regents exams and found a significant tendency for local school districts to award full credit on questions requiring scorer judgment even when the exam answers were vague, incomplete, inaccurate or insufficiently detailed. As a result, scores awarded by the local school districts often were higher than the scores determined by the expert review team. Despite the seriousness of the review team findings and questions raised about the accuracy and reliability of Regents exam scoring, there was little evidence that SED took action to follow up to address these matters with the officials of local school districts where the variant scoring took place.

We further found that when local school districts fail to comply with SED requests to submit scored exams for further review, SED was not following up to obtain these examinations. We recommend follow up take place because there is considerable risk that the failure to submit scored exams may be a willful attempt to avoid scrutiny of scoring accuracy. We also concluded that SED

had limited assurance that exam raters actually attended annual training for scoring exams. We recommend that SED strengthen its formal guidance for administering the exams to better ensure that raters attend such training.

In their response to our draft audit report, Department officials agreed with 11 of the report's 12 recommendations. Officials indicated that they recognize the need to strengthen oversight of local scoring of Regents Examinations, and they are implementing additional procedures to expand their monitoring efforts.

This report, dated November 19, 2009, is available on our website at: <http://www.osc.state.ny.us>.

Add or update your mailing list address by contacting us at: (518) 474-3271 or

Office of the State Comptroller

Division of State Government Accountability

110 State Street, 11th Floor

Albany, NY 12236

Introduction

Background

Regents exams are statewide tests that are given each year in particular subject areas, such as English, history, mathematics, science, and foreign languages. They are intended to assist colleges in making admission decisions and provide measures of students' academic performance and schools' effectiveness and adherence to the State's prescribed curricula. In addition, beginning in the 1990s, most high school students in the State were required to pass certain Regents exams in order to earn a high school diploma.

Within the State Education Department (SED), the Office of State Assessment oversees the development and administration of the Regents exams. Local school officials are responsible for scoring the exams and reporting the results to SED.

SED issues a Scoring Key and Rating Guide (also known as Rubrics) for each Regents exam. The Rubrics contain the correct answers for the questions with one correct answer (e.g., multiple choice questions), and examples of acceptable answers for the questions with more than one acceptable answer (e.g., fill-in-the-blank questions and essays). The Rubrics also contain guidelines for awarding partial credit where applicable (such as on essays), and instructions for converting the "raw" exam score to the final published score.

SED also has specialists in each examination subject, who may be consulted by the schools during the examination period if it is not clear how a particular question or exam should be scored. To further ensure scoring accuracy and consistency throughout the State, SED hires expert consultants to analyze scoring variations among schools and individual raters and to periodically perform a statewide review to assess the accuracy of local scoring practices.

Regents exams are scored (or "rated") by teachers at the schools giving the exams. The school principal is responsible for selecting the raters for each exam and monitoring the scoring process to ensure that it is performed in accordance with SED guidelines. Usually, the rater is a teacher who is responsible for teaching the subject covered by the exam.

According to SED guidelines, all raters must be thoroughly familiar with the rating instructions for their exams. Training sessions in the scoring process are provided each year by local Boards of Cooperative Educational Services (BOCES) and other designated trainers, and the raters may attend these sessions. The raters may also attend in-house training sessions at their

schools. The raters and other school administrators involved in the scoring process are required to sign a certification stating that they have followed the rules for administering, supervising and scoring the exams.

Regents exams are given statewide in June, August and January. The exam results are included in the annual report cards the State publishes for each school district, and are taken into account when the academic performance of the districts is evaluated.

Audit Scope and Methodology

We audited selected aspects of SED's oversight of the scoring of Regents exams for the period July 1, 2006 through March 31, 2009. To accomplish our audit objective, we interviewed officials at SED and selected school districts, examined relevant SED policies and procedures, reviewed documents and reports prepared by and for SED, and reviewed applicable sections of State laws and regulations.

We conducted our audit in accordance with generally accepted government auditing standards. 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 based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objective.

In addition to being the State Auditor, the Comptroller performs certain other constitutionally and statutorily mandated duties as the chief fiscal officer of New York State. These include operating the State's accounting system; preparing the State's financial statements; and approving State contracts, refunds, and other payments. In addition, the Comptroller appoints members to certain boards, commissions and public authorities, some of whom have minority voting rights. These duties may be considered management functions for purposes of evaluating organizational independence under generally accepted government auditing standards. In our opinion, these functions do not affect our ability to conduct independent audits of program performance.

Authority

We performed this audit pursuant to the State Comptroller's authority as set forth in Article V, Section 1 of the State Constitution and Article II, Section 8 of the State Finance Law.

We provided a draft copy of this report to Department officials for their review and formal comment. We considered the Department's comments in preparing this report. Department officials agreed with 11 of our report's 12 recommendations and indicated that they recognize the need to strengthen oversight of local scoring of Regents Examinations. Consequently, officials are implementing additional procedures to expand their monitoring efforts.

With regard to Regents Examinations (obtained from school districts) that were re-scored by Department personnel, officials indicated that the overall rates of agreement (reliability) were statistically high, although agreement rates for certain essay questions were comparatively low. The Department's comments are included in their entirety at the end of this report. Our rejoinders to the Department's comments are included thereafter in our State Comptroller's comments.

Within 90 days of the final release of this report, as required by Section 170 of the Executive Law, the Commissioner of the State Education Department shall report to the Governor, the State Comptroller, and the leaders of the Legislature and fiscal committees, advising what steps were taken to implement the recommendations contained herein, and where recommendations were not implemented, the reasons therefor.

**Contributors
to the Report**

Major contributors to this report include Steven Sossei, Brian Mason, William Clynes, Mary Roylance, Laurie Burns, Andrea Dagastine and Dana Newhouse.

Audit Findings and Recommendations

SED Oversight

Regents examinations are intended to provide reliable measurements of both individual student and overall school academic performance. If the exams are to effectively serve these purposes, they must be scored accurately and consistently throughout the State. SED's periodic statewide reviews of the schools' scoring practices (called Department Reviews) are intended to assess the accuracy and consistency of the scoring process. Accordingly, our report focuses on these reviews.

In a Department Review, a group of experienced high school teachers, led by SED's subject specialists, re-scores a sample of Regents examinations, compares its scores to the original scores, and assesses the accuracy of the original scores. The Review team produces a final report for SED summarizing the results of its review, and may make recommendations to SED for improving the accuracy and consistency of the scoring process. SED then writes to the schools included in the sample to inform them of the results for their exams.

Department Reviews are not performed for every examination period. The Reviews that are performed cover a single examination period (e.g., June) and certain of the examinations given during that period. SED selects a random sample of schools for each exam to be reviewed, and asks the schools to send all their examination papers in that subject to the Review team. SED may also include additional schools in the sample based on past observations or complaints about the scoring practices at those schools. The Review team then selects certain of the examination papers for review. During the review process, the Review team focuses on the questions with more than one acceptable answer (e.g., fill-in-the-blank questions and essays). It re-scores those questions, and compares its scores to the original raters' scores. When there are discrepancies, the school principals are asked to review the questions and make any adjustments they believe are appropriate to the students' exam scores. However, any such adjustments are made independently of the review process, are not reported to SED, and are not further considered by the Review team as part of its assessment of the accuracy of scoring.

At the time of our audit, the most recently completed Department Review covered the exams that were given in June 2005. Two exams were selected for review, a sample of about 200 schools was selected for each exam, and more than 5,600 individual examination papers were re-scored (2,393 for Exam A and 3,209 for Exam B). The Review team found that, for the 22 questions that were re-scored on the two exams (9 in Exam A and 13 in

Exam B), the scores awarded by the schools were consistently higher than the scores awarded by the Review team, as follows:

- For Exam B, the schools' total raw scores on the 13 re-scored questions were higher than the total scores awarded by the Review team on 80 percent of the examination papers reviewed. The total raw scores were the same on 15 percent of the papers, and the school's total raw scores were lower on the remaining 5 percent.
- For Exam A, the schools' total raw scores on the nine re-scored questions were higher than the total scores awarded by the Review team on 58 percent of the examination papers reviewed. The total raw scores were the same on 32 percent of the papers, and the school's total raw scores were lower on the remaining 10 percent.
- On Exam B, the schools' total raw scores on the 13 re-scored questions were at least three points higher (or lower) than the total scores awarded by the Review team in 34 percent of the examination papers reviewed. Since a difference of one point in the raw score can result in a difference of several points in the final score, a difference of three points in the raw score can result in a difference of ten or more points in the final score.
- On Exam A, the schools' total raw scores on the nine re-scored questions were at least three points higher (or lower) than the total scores awarded by the Review team in 17 percent of the examination papers reviewed.
- Exam B contained two five-point essay questions. On these two questions, the schools' raw scores were higher (or lower) than the Review team's raw scores in 47 percent and 43 percent, respectively, of the examination papers reviewed.

In its final report, the Review team noted that the schools tended to award full credit even when answers were vague, incomplete, inaccurate or insufficiently detailed, and as a result, their scores tended to be higher than the scores awarded by the Review team. The Review team recommended that improvements be made in scoring training and built-in quality control during the scoring process (e.g., to guard against the effects of rater fatigue, some states require that each rater's exams be periodically reviewed by another scorer during the scoring process).

We reviewed the actions taken by SED in response to the results of the 2005 Department Review. Despite the seriousness of the Review team's findings and the questions they raised about the accuracy and reliability of Regents examination scoring process, we found little evidence action had been taken by SED to address the scoring weaknesses identified by the Review team.

For example, we found no evidence actions were taken to implement the Review team's recommendations to improve scoring training and enhance quality control during the scoring process. We also found no evidence actions were taken to bring about improvements at particular schools. While SED wrote to the schools selected in the re-scoring samples to inform them of the results for their exams, SED required no further actions from the schools, even when there were significant scoring discrepancies on their exams. Rather, SED informed the schools that the Department Review is intended to be used as a training tool and to provide schools with useful information.

We recognize that the scoring discrepancies on a particular school's exams could fall within an acceptable range of error/variation. Therefore, to make it clear to school officials when improvements are needed in their scoring practices, we recommend SED establish an acceptable range for scoring discrepancies for each exam reviewed. Then, SED can evaluate each school's performance on the basis of this criteria (e.g., "acceptable" or "unacceptable"), and report more meaningful evaluation results to the schools. We further recommend that SED require the schools with "unacceptable" discrepancies to develop corrective action plans, and follow up with these schools to determine whether the plans are being implemented.

We note that an earlier consultant's analysis identified the same scoring weaknesses as the 2005 Department Review. This analysis covered the 2003-2004 school year and was performed for SED by CTB/McGraw-Hill. In this analysis, a sample of Regents examinations was re-scored and the consultant found that its scores were generally lower than the scores awarded by the schools. Similar to the 2005 Review team, the consultant noted that improvements in scoring training and supervision would likely increase the reliability and validity of the Regents examination scoring process.

In addition, we determined that SED's oversight of scoring practices would be strengthened if the following improvements were made in the Department Reviews:

- In the Reviews, SED selects a sample of schools for each exam that is to be re-scored, and asks the schools to send all their exams in that subject to the Review team. However, we found that some schools do not comply with this request. For example, in the 2005 Department Review, 18 of the 192 selected schools did not provide the requested examination papers for Exam A, and 20 of the 205 selected schools did not provide the requested examination papers for Exam B. SED officials told us they follow up with such schools and repeat their request, but often proceed without these exams because of the tight time schedule for the review process. We acknowledge the need for a tight time schedule,

but recommend SED obtain and review all requested examination papers, even if the papers cannot be included in that particular Department Review, because there is a considerable risk some examination papers might be deliberately withheld to avoid scrutiny.

- The sampling process for the Reviews is random, with some schools added to the sample because they are believed to be at risk for scoring irregularities. However, this approach does not ensure that all schools with a significant presence in the Regents examination program are selected for review within a reasonable period of time. We recommend the random selection process be modified to provide such assurance. We also recommend that a formal risk assessment process be used to assign risk to the schools. The current process is informal, and as a result, there is less assurance risk is properly assessed.
- At the time of our audit, the most recently completed Department Review covered the exams that were given in June 2005. SED officials told us that the next Department Review covered exams given in January 2008. However, at the time of our audit, the Office of State Assessment had not completed it. SED officials also told us that they planned to conduct a Department Review of exams given in January 2009. Nonetheless, we question the adequacy of SED's oversight given the results of the June 2005 Review and the absence of any completed Reviews since that time. Consequently, we recommend that SED initiate and complete Reviews annually.

Also, SED officials told us they investigate all complaints about Regents examination scoring practices outside New York City. (The New York City Department of Education is responsible for investigating such complaints within the City.) According to officials, BOCES district superintendents represent SED in the field and conduct the investigations at SED's request. SED officials provided us with a log of 13 complaints made in recent years. However, we found evidence of only five investigations, of which only one corresponded to the complaint log provide by SED. Thus, we concluded that the complaint log was incomplete, and there were no investigation reports for 12 of the 13 complaints on the log. Officials told us that additional investigations were conducted, but documentation of the investigations was not maintained. We recommend that officials update the complaint log timely and accurately, and maintain adequate documentation of investigations of complaints. In the absence of this documentation, there is little assurance that complaints are thoroughly investigated and properly resolved.

In response to our audit findings, SED officials stated that they have limited financial and human resources to address the accuracy of Regents exam scoring, and they have decided to allocate their limited resources to other

responsibilities, such as developing exam questions and scoring keys. Officials also said they are concerned about the accuracy of scoring and have explored various options, such as third-party scoring, but they lack the necessary resources to pursue these options at this time.

SED officials further noted that, in 2006, they contracted with a consulting firm to prepare technical reports addressing various aspects of Regents exam scoring. They noted that the data collected for these reports could be used in analyses that would identify schools with potential scoring problems. Officials also told us they started a data analysis project in 2007, but staffing issues have limited the project's development.

We acknowledge that many demands are made on SED in its administration of Regents exams. However, the integrity of the exams must also be a priority, and therefore, SED must adequately oversee local scoring practices. In its Department Reviews, SED has developed an excellent means of providing such oversight. However, SED is not making effective use of this monitoring tool, because it is not following through with corrective actions to address the questionable scoring practices that have been identified. We recommend SED take such actions.

- Recommendations**
1. Implement the improvements recommended by the 2003-04 consultant and 2005 Review team reports, or take alternative actions to address the questionable scoring practices they identified.
 2. For each Regents exam that is re-scored in a Department Review, establish an acceptable range for the scoring discrepancies between the Review team and the original raters. Evaluate each school in the sample on the basis of the criteria, and report the evaluations to the schools.
 3. Request schools with significant exam scoring deficiencies to advise SED of any changes made to exam scores as a result of errors identified by the Department Review.
 4. If a school's scoring practices are found to be unacceptable, require the school to develop a corrective action plan and follow up with the school to determine whether the plan is being implemented.
 5. Obtain and review all examinations that are requested from schools during a Department Review, even if the papers cannot be included in that particular Review.
 6. Modify the process for selecting schools in Department Reviews to ensure that all schools with a significant presence in the Regents examination program are selected for review within a reasonable period of time.

7. Develop a formal process of assessing a school's risk for irregularities in its scoring of Regents examinations, use this process to assign a level of risk to all the schools in the Regents examination program, and routinely include a certain number of high-risk schools in each Department Review.
8. Perform Department Reviews annually.
9. Expedite the completion of the January 2008 Department Review.
10. Ensure that the examination complaint log is kept up-to-date and accurate, and maintain documentation of all investigations of complaints about examination scoring practices outside New York City.

Training in Examination Scoring

The scoring process for Regents exams can often be complex. The Rubrics for each exam are several pages long and often contain detailed scoring guidance for questions, particularly those that require a student to fill-in-the-blank, solve a problem, or write an essay. Further, according to Department guidelines, all teachers involved in rating Regents exams must be thoroughly familiar with the Department's rating instructions to maintain uniform rating standards. Therefore, SED officials advised us that all exam raters should receive annual training in the Regents exam scoring process. Training sessions in exam scoring are provided annually by local BOCES or other SED-designated trainers. Training sessions are also conducted at individual schools by district personnel using SED guidance materials. According to SED officials, exam raters should receive training in exam scoring annually (even if they have received such training sometime in the past) because there can be significant changes in the Rubrics from one year to the next.

We determined, however, that SED has little assurance that exam raters actually receive scoring training annually. Without such training, raters might not be adequately prepared to score exams, and the potential of scoring errors could be increased. Further, we believe there is considerable risk that deficiencies in rater training contributed to the questionable scoring practices that were identified in the 2005 Department Review and other reviews of exam scoring accuracy. As noted previously, the 2005 Review recommended that the Department make improvements in exam scoring training - which could include ensuring that raters attend the annual training.

Based on the results of our review, we recommend that SED amend its formal guidance pertaining to raters' participation in training for exam scoring. At the time of our review, SED did not require schools to document raters' attendance at annual training sessions for exam scoring. Consequently, the Department (and district officials as well) had limited assurance that all

raters actually attended such training. In addition, SED should modify the certifications (that exam raters must sign) to affirm compliance with SED guidance for administering Regents exams. Currently, such certifications include statements pertaining to the supervision and grading of the exams. However, the certifications do not address raters' participation in the training for scoring them. We concluded that raters should certify that they have participated in the training.

- Recommendations**
11. Advise school districts to maintain documentation of their raters' attendance at training sessions for Regents exam scoring.
 12. Expand the formal rater exam certification to include an affirmation that the rater attended training for exam scoring.

Agency Comments



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

PRESIDENT OF THE UNIVERSITY
AND COMMISSIONER OF EDUCATION

September 14, 2009

Mr. Brian E. Mason
Audit Director
Office of the State Comptroller
Bureau of State Audit
110 State Street – 11th Floor
Albany, NY 12230

Dear Mr. Mason:

The following is the New York State Education Department's (Department) response to the Office of the State Comptroller's draft audit report (2008-S-151) of the State Education Department, Oversight of Scoring Practices on Regents Examinations. The recommendations are helpful and implementation will enhance our processes to ensure accurate scoring of Regents Examinations. I have provided additional information in response to the report's observations on Department oversight and a response to each of the recommendations.

While the audit report emphasizes discrepancies in scoring between the schools and the State Education Department's scorers, there is additional important information about scoring accuracy that presents a more balanced picture of the facts. Overall, the rate of agreement was statistically high. When evaluating whether the schools scored the exams accurately, it is essential to keep this in mind:

- The rate of agreement (or reliability) between the schools and the State Education Department's scorers was very high statistically overall – a total of .92 – on one of the two exams selected for rescoring in 2005.
- The exact agreement was lower but still good on the second exam; the schools' teachers and the Department's rescorsers exactly agreed 83% of the time. Although exact agreement rates were low for the two 5-point essay items (53% for one and 57% for another), when the adjacent agreement is calculated (adjacent agreement means a score difference of +1/-1 point), the total agreement for the two 5-point questions increased to 92% and 96% respectively.
- Since the audit was conducted, we have completed the review of the Regents English Examination for 2008. Again, the correlation on scoring was high – a total of .89.

*
Comment
1

*
Comment
2

* See State Comptroller's Comments on page 27.

In order to help ensure that Regents Examinations are scored properly, we currently take the following actions:

- Conduct monitoring visits by the Department, BOCES administrative, and large-city school district administrative staff to approximately 40 schools during the administration of Regents Examinations;
- Establish and maintain a computerized database for recording and follow up to all allegations of improper conduct in the administration and scoring of all State examinations, including Regents Examinations; and
- Take action in confirmed and/or suspicious cases including requiring schools to conduct their scoring of Regents Examinations at an alternate site such as a BOCES and/or prohibiting specific individuals from participation in the administration and/or scoring of Regents Examinations.

To improve our oversight of the scoring process, we will also take the following actions in the future:

- Annual Department reviews of selected Regents Examinations, with schools chosen through a risk analysis;
- More aggressive follow-up whenever problems are noted to ensure that scoring procedures are corrected in the future;
- Require schools to undertake a corrective action plan when warranted; and
- Require all teachers who score Regents Examinations to attest that they have passed the appropriate training sessions.

We have recognized the need to strengthen oversight of local scoring of Regents Examinations and are implementing additional procedures to expand our monitoring. The processes we will implement are:

- Conducting Regents Examination Item-Level Score Collections at both the representative sampling and full test taking census levels;
- Conducting erasure and response tampering analysis of Regents Examination answer papers at the full test taking census, representative sampling, and selected schools levels;
- Conducting spike/cluster analysis of Regents Examination scores to identify schools with suspicious results to be conducted annually; and
- Using combinations of the processes listed above in tandem with one another to more effectively identify concerns with local scoring of Regents Examinations and/or to prevent reoccurrences of improper conduct in the administration and scoring of them.

Below you will find our responses to the recommendations contained in the draft audit report.

Recommendation 1:

Implement the improvements recommended by the 2003-04 consultant and 2005 Review team reports, or take alternative actions to address the questionable scoring practices they identified.

We agree with this recommendation. As we proceed to explore different alternatives to address scoring practices, we will continue to consider the recommendations resulting from the Department's review process. As an alternative action, the Department has implemented a statewide turnkey train-the-trainer approach whenever new types of constructed-response questions are introduced into new or pre-existing Regents Examinations. The Department most recently conducted such training to enable teachers to reliably administer and score new types of questions used in a new Earth Science Performance Test. Each teacher/scoring leader who participated in one of the Department's regional train-the-trainer sessions then conducted one or more scorer-training sessions in his/her region.

For teachers who are new to scoring established examination questions, scorer training is most often conducted by the BOCES and large city school districts. The Department will work with these entities to ensure they provide us with records of the new scorer training sessions which they conduct.

The Department recognizes the tendency noted in the audit report for local district scores to be higher when discrepancies are found between the scores awarded by local school districts and Department Review team scores. In addition to Department Reviews, we will use a new variety of auditing processes to identify schools that exhibit unacceptable frequencies of discrepant scoring. The Department will require remedial actions for the scoring of examination papers from these schools. (See Recommendation 4)

Recommendation 2:

For each Regents exam that is re-scored in a Department Review, establish an acceptable range for the scoring discrepancies between the Review team and the original raters. Evaluate each school in the sample on the basis of the criteria, and report the evaluations to the schools.

We agree. When conducting a variety of examination score monitoring procedures, including Department Reviews, the Department employs multiple statistical methods to assess the scoring reliability of the Regents Examinations. As a result, we are able to identify schools with scoring that is most discrepant from State scoring, as well as schools with scoring that is most consistent. The Department will provide feedback to schools at both ends of that continuum.

Recommendation 3:

Request schools with significant exam scoring deficiencies to advise SED of any changes made to exam scores as a result of errors identified by the Department Review.

We disagree. Reviews of Regents Examination scoring are conducted on a sampling basis after the tests are given. The purpose is to determine the reliability of scoring and, when applicable, to suggest areas for future improvement. It is not feasible to require schools to change only the scores of the sampled tests; by the time the scores are collected and analyzed, the students have moved on in their academic program.

Recommendation 4:

If a school's scoring practices are found to be unacceptable, require the school to develop a corrective action plan and follow up with the school to determine whether the plan is being implemented.

We agree. To the extent that staffing permits, the Department will require schools with unacceptable scoring practices to submit corrective action plans and will follow up with those schools to ensure that the plans are being implemented. In the most extreme cases, we will mandate the corrective actions that a school must implement, such as a specific training or scoring at an alternate site with increased monitoring for a specified period of time.

Recommendation 5:

Obtain and review all examinations that are requested from schools during a Department Review, even if the papers cannot be included in that particular Review.

We agree. The Department will continue to follow up with all schools that do not promptly submit requested answer papers, examination and scoring certificates, or other requested examination related materials to ensure that all documents are received and reviewed by the Department.

Recommendation 6:

Modify the process for selecting schools in Department Reviews to ensure that all schools with a significant presence in the Regents examination program are selected for review within a reasonable period of time.

We agree. All schools with significant presence in the Regents Examination program will be included in a variety of Department-conducted monitoring.

Recommendation 7:

Develop a formal process of assessing a school's risk for irregularities in its scoring of Regents examinations, use this process to assign a level of risk to all the schools in the Regents examination program, and routinely include a certain number of high-risk schools in each Department Review.

We agree. When conducting Department Reviews, score collections, and other monitoring processes, the Department will identify high risk schools. This will be done by using statistical methods to identify schools with scoring of Regents Examinations that is most discrepant from State scoring. These schools will be subject to increased inclusion in future score monitoring processes including Department Reviews. In extreme cases, schools may be required to conduct their scoring of examinations at an alternate site such as a BOCES.

Recommendation 8:

Perform Department Reviews annually.

We agree. The Department will attempt to conduct at least one Department Review annually as it has done for the past two school years. In tandem with Department Reviews, we will also increasingly utilize other exam score monitoring activities that it has researched, including score collection and data analysis. These tools provide an excellent vehicle for following up on concerns that are noted in Department Reviews and for identifying specific schools and examination titles to include in future Department Reviews.

Recommendation 9:

Expedite the completion of the January 2008 Department Review.

This has been done. We are currently working on finalizing the technical reports for the January 2009 Department Review and for the June 2009 score collections for the selected Regents Examinations. It should be noted that, except for the completion of the technical report, the January 2008 Department Review was completed before the audit began.

Recommendation 10:

Ensure that the examination complaint log is kept up-to-date and accurate, and maintain documentation of all investigations of complaints about examination scoring practices outside New York City.

We agree. The Department already maintains a database and updates it continuously to ensure it is accurate.

Recommendation 11:

Advise school districts to maintain documentation of their raters' attendance at training sessions for Regents exam scoring.

We agree. The Department will collaborate with the district superintendents, large city assessment directors, and professional organizations on better ways to maintain records of teacher attendance at training sessions for Regents Examination scoring.

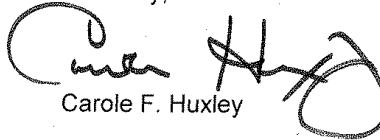
Recommendation 12:

Expand the formal rater exam certification to include an affirmation that the rater attended training for exam scoring.

We agree. To achieve this, the Department will develop and implement a teacher scoring certificate. On this certificate, all teachers will attest to having participated in a training session specific to the scoring of the constructed-response questions on a specific Regents Examination before they began scoring actual student responses. All participating schools will be required to complete this certificate and return it to the Department. Schools that fail to return the certificate with appropriate signatures will be contacted.

If you have any questions regarding this response, please contact Johanna Duncan-Poitier, Senior Deputy Commissioner of Education – P-16, at (518) 474-3862.

Sincerely,


Carole F. Huxley

c: Johanna Duncan-Poitier

State Comptroller's Comments

1. We acknowledge that the overall statistical correlation between school district exam scorers and SED re-scorers was relatively high (.92) for the exam in question. However, this does not mean that 92 percent of the exams reviewed by SED were scored correctly by the districts. In fact, for one 3-point question on this exam, SED re-scorers disagreed with district scorers 35 percent of the time. In most of these instances, students received one or two points more for this question than SED would have awarded. As noted in our report, a difference of one point in the raw score of an exam paper can correspond to a difference of several points in the final score - which can be the difference between passing and failing the exam.
2. We acknowledge that, in the aggregate, district exam scorers and SED re-scorers agreed 83 percent of the time for the exam in question. However, the overall statistical correlation between the districts' and the SED's scoring for this exam was only .71. Moreover, this does not mean that the exams in question were scored correctly 83 or 71 percent of the time. In fact, for three constructed response questions on this exam, SED re-scorers disagreed with district scorers 36 percent or more of the time. In most of these instances, students received more credit than SED would have awarded.