New York State Examination Programs
Department Review of January 2009
Regents Competency Tests in Reading and Science

Technical Report

October 2009

Office of State Assessment
New York State Education Department

Copyright

Table of Contents

List of Tables

Section I: Introduction

This report summarizes the results of a Department Review of the Regents Competency Tests (RCT) in Reading and in Science administered in January 2009. Department Review is an internal audit process conducted by the New York State Education Department (the Department) to ensure the reliability of the state's high school level examination program. Each year, to ensure the reliability of local scoring of examinations, the Department conducts audits of New York State teachers' local scoring of a selected number of high school level Regents Examinations and RCTs. In the 2008-2009 school year, the January RCTs in Reading and Science were chosen for Department Review. The 2009 Department Review focused on the rescoring of all items for both exams. Student test papers from a sample of schools from across the state were collected and answer sheets for each test were rescored by the State's independent scorers.

The purpose of the rescoring is to provide necessary test reliability and inter-rater reliability evidence for the high school level examinations. The audit process also allows the Department to evaluate the extent to which teachers and committees of teachers are properly applying the scoring rubrics and scoring guides when scoring their tests. Department Review also acts as an added incentive to schools and teachers, ensuring that they score tests properly in accordance with overall state directions and oversight. The process also provides feedback to schools, which can lead them to improving their scoring procedures and enhance compliance with the scoring rubrics if deficiencies are noted. The process of Department Review is an important element for maintaining overall test reliability.

Section II: Review Procedure

Sample Collection and School Participation

A stratified random sample of 120 high schools was selected for the January 2009 Department Review, sixty for each test. No school was chosen for both tests. The school sample was stratified by Need/Resource Capacity Category to represent the New York State school population (see Table 1).

Table 1: Need/Resource Category (NRC) Definitions

Need/Resource Category	**Definition**
New York City	New York City
Big 4 Cities	Buffalo, Rochester, Syracuse, Yonkers
High Need Urban/Suburban	Districts at or above 70th percentile on the index with at least 100 students per square mile or enrollment greater than 2500
High Need Rural	All districts at or above the 70th percentile with fewer than 50 students per square mile or enrollment of less than 2500
Average Need	All districts between the 20th and 70th percentiles on the index
Low Need	All districts below the 20th percentile on the index
Charter Schools	Each charter school is a district

Of the 60 schools selected for the RCT in Reading, 42 returned examination papers to the Department. (The Department also followed up with the 18 schools that did not submit papers.) Of the 42 schools returning examination papers, 3 sent non-scored exams. Upon receipt of the examination papers from the sample schools, a random sample of up to 10 papers of the obtained examination papers from each school was selected for rescoring by an independent group of raters. Again, for schools with 10 or fewer examination papers, all papers were rescored. A total of 366 examination papers were rescored by the Department's independent raters.

Of the 60 schools selected for the RCT in Science, 34 returned examination papers to the Department. (The Department also followed up with the 26 schools that did not submit papers.). Of the 34 schools returning examination papers, 5 sent partially-scored or non-scored exams. Upon receipt of the examination papers from the sample schools, a random sample of up to 10 papers of the obtained examination papers from each school was selected for rescoring by an independent group of raters. For schools with 10 or fewer examination papers, all papers were rescored. A total of 288 examination papers were rescored by the Department's independent raters.

Section III: Data Analysis

Data Preparation

An Excel spreadsheet was prepared for individual student data entry. Both the final local school scores and the state rescores were entered and saved for data analysis. A total of 366 records was entered for RCT in Reading. A total of 288 records was entered for RCT in Science.

Response data were obtained from two sources. Each student had one score from local scoring and one score from state scoring. Student local scores and state audit scores were matched for data analysis. The matching local score and state rescore is 100%. Only records with matching data for both local and state scoring were used in data analysis.

Table 2: Number of Records Received

	Number of Records	
	Local	**State**
RCT in Reading	366	366
RCT in Science	288	288

Additonal Analysis

School level erasure marks were also prepared for analysis and entered into the spreadsheet file. Each record contains school name, BEDS code, total papers received, total erasures and total changes from wrong to right.

Methods Used

Multiple methods were employed to assess the scoring reliability of the January 2009 RCTs in Reading and Science. The following methods address the degree of agreement between local school scores and state rescores.

1. **Test Mean and Standard Deviation:** Test raw score mean difference and standard deviation between the local school scores and state rescores were calculated as measures of average agreement/difference and variability between the two groups of scorers on a given answer paper.

2. **Inter-rater Percent Agreement:** Raw score agreement, as a measure of consensus between local school scorers and state rescorers, was calculated for each test. In this method, the percentage of exact agreement (i.e., local scores match state rescores) and the percentage of adjacent and nonadjacent agreement (local scores and state scores differ in their score assignment by 1, 2 to 5 or more score points) were calculated.
3. **Total Score Correlation:** A local total score and state total score based on raw score results were calculated. Correlation between the two total scores was calculated to provide an overall measure of the scoring reliability.
4. **Erasure Analysis:** Student physical papers were examined to count the total erasures. Focus is on the school level summary counts of changes from wrong to right. Mean erasures were calculated and exact binomial distributions were used to estimate the probabilities of the extremeness of the counts.

Section IV: Results

Raw Score Mean and Standard Deviation

Test raw score analysis was performed on both the RCT in Reading and the RCT in Science. Table 3 presents the comparison of local and state raw score mean and standard deviation. Individual student score mean and standard deviation are measures of average agreement/difference and variability between the two groups of scorers. The results show very close agreement between local and state total test raw scores.

Test score ranges from a minimum of 8 to a maximum of 69 for the RCT in Reading and 11 to 64 for the RCT in Science. Differences in standard deviation between local and state scoring were minimal.

Table 3: Comparison of Total Mean Test Score and Standard Deviation

		Mean			Standard Deviation		
	N-Count	**Local**	**State**	**Difference**	**Local**	**State**	**Difference**
RCT in Reading	366	41.67	41.68	0.01	12.85	12.79	-0.06
RCT in Science	288	36.67	36.67	0.00	10.21	10.18	-0.03

Inter-rater Percent Agreement

Inter-rater agreement was conducted to measure the difference between local scoring and state rescoring. The percentage of times local scores and state rescores agreed and differed was calculated. Table 4 shows that the exact agreement between local and state score ranges is 94.5% for the RCT in Reading and 93.1% for the RCT in Science. The total agreement was 99.4% for the RCT in Reading and 99.8% for the RCT in Science, respectively. Table 5 presents the percentage of raw score differences.

Table 4: Inter-rater Percent Agreement between State and Local Scores

Item	Max Points	N-Count	Agreement (%)		
			Exact Agreement	Adjacent Agreement (+/- 1 Credit)	Near Adjacent Agreement (+/- 2 to 3 Credits)
RCT in Reading	70	366	94.5	3.3	1.3
RCT in Science	70	288	93.1	5.2	0.9

Table 5: Percentage of Raw Score Differences

Item	Max Points	N-Count	Percentage of Score Difference						
			0	1	2	3	4	5	6 or more
RCT in Reading	70	366	94.5	3.3	1.3	0	0	0.3	0.6
RCT in Science	70	288	93.1	5.2	0.3	0.6	0.4	0.3	0

Total Score Correlation

As an overall measure of scoring reliability, the Pearson Correlation Coefficient between the local and state total raw scores was computed. This statistic is often used as an overall indicator of scoring reliability and generally ranges from 0.00 to near 1.00. The correlation coefficient between the local and state total test scores for the RCT in Reading was 0.99, which indicates a high degree of scoring reliability. The correlation coefficient between the local and state total test scores for the RCT in Science was also 0.99, a high degree of scoring reliability.

Erasure Analysis

Student physical papers were examined to count the total erasures. School level counts of erasures were summarized for a total of 39 schools who submitted papers for the RCT in Reading. The percent of total school erasures were obtained using total possible erasures as a denominator. Mean percentage obtained was 4%. There is only one school that had the percentage go above 2 standard deviations. (The Department followed up with the identified school). As to the RCT in Science, the mean percentage was 3%. One school with percentage above 2 standard deviations was also identified. (The Department also followed up with the identified school.) Overall, the erasure analysis results indicate a standardized test administration.

Section V: Summary

The Department Review is an internal audit process to ensure the scoring reliability of New York State high school level examinations. The January 2009 RCTs in Reading and Science were chosen for Department Review. A sample of 39 schools for the RCT in Reading and a sample of 29 schools for the RCT in Science submitted their January 2009 operational test papers to the Department for rescoring by an independent group of scorers.

A total of 654 test papers from schools across New York State were rescored. Multiple statistical methods were employed to assess the scoring reliability of the total test raw score. A comparison based on average means and standard deviations showed a very close agreement between local scoring and state rescoring. The inter-rater agreement between local scoring and state rescoring also indicated a high degree of agreement. The exact agreement between local and state scoring was 94.5% for the RCT in Reading and 93.1% for the RCT in Science.

As an overall measure of scoring reliability, the Pearson Correlation Coefficient between the total local scores and total state scoring was .99, which also indicated a high degree of scoring reliability.

In general, the Department Review has found a high degree of agreement between local scoring and state rescoring. Erasure analysis of operation test papers using percentage norm, indicates a proper test administration.