

**Inter-Rater Reliability Study of New York State
Grades 3–8 Mathematics Test
2009 Test Administration**



Technical Report

Prepared for the New York State Education Department
by Pearson

September 29, 2009

Table of Contents

I.	Purpose and Scope of Audit	1
	Purpose	1
	Scope	1
II.	Selection of School Sample and Test Booklets	3
	Audit Samples	3
	Stratified Sampling Design at the School Level	5
III.	Data Collection and School Participation	7
IV.	Selection and Training of Auditors	8
	Description of How the Auditors Were Selected	8
	Training of Auditors	8
	Quality Control Procedures	9
V.	Audit Procedures	10
	Description of the Audit Procedures	10
VI.	Data Analysis	11
	Item Means	11
	Intra-Class Correlation	11
	Weighted Kappa	12
	Inter-Rater Agreement	12
	Total Score Correlation	12
VII.	Results	13
	Item Means	13
	Percent of Agreement	13
	Intra-Class Correlations	13
	Weighted Kappa	17
	Inter-Rater Agreement	17
	Total Score Correlation	20
	Additional Analyses	21
VIII.	Summary	22
	References	22
	Appendix A	23
	Appendix B	26
	Appendix C	30
	Appendix D	48
	Appendix E	66
	Appendix F	70
	Appendix G	84
	Appendix H	98

List of Tables

Table 1.	Need/Resource Capacity Category Definitions	3
Table 2.	State N-counts.....	4
Table 3.	Target Proportions.....	4
Table 4.	Target N-counts.....	5
Table 5.	Selected N-counts	5
Table 6.	Sample Proportions	6
Table 7.	Obtained N-counts for Mathematics	7
Table 8.	Obtained Proportions for Mathematics	7
Table 9.	New York State Public Schools Mathematics Operational Test 2009: Inter-Rater Agreement	14
Table 10.	Percentage of Absolute Raw Score Difference between Local and Audit Scores.....	18
Table 11.	Correlations Between Local and Audit Scores	20
Table C-1.	NYS Public Schools Grade 3 Mathematics Operational Test 2009: Inter-Rater Agreement	31
Table C-2.	NYS Public Schools Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement	33
Table C-3.	NYS Public Schools Grade 5 Mathematics Operational Test 2009: Inter-Rater Agreement	37
Table C-4.	NYS Public Schools Grade 6 Mathematics Operational Test 2009: Inter-Rater Agreement	39
Table C-5.	NYS Public Schools Grade 7 Mathematics Operational Test 2009: Inter-Rater Agreement	42
Table C-6.	NYS Public Schools Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement	44
Table D-1.	NYS Public Schools (Without NYC) Grade 3 Mathematics Operational Test 2009: Inter-Rater Agreement	49
Table D-2.	NYS Public Schools (Without NYC) Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement	51
Table D-3.	NYS Public Schools (Without NYC) Grade 5 Mathematics Operational Test 2009: Inter-Rater Agreement	55
Table D-4.	NYS Public Schools (Without NYC) Grade 6 Mathematics Operational Test 2009: Inter-Rater Agreement	57
Table D-5.	NYS Public Schools (Without NYC) Grade 7 Mathematics Operational Test 2009: Inter-Rater Agreement	60
Table D-6.	NYS Public Schools (Without NYC) Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement	62
Table E-1.	NYC Public Schools Grades 3–8 Mathematics Operational Test 2009: Inter-Rater Agreement	67

List of Tables (continued)

Table F-1.	NYS Public Schools Grade 3 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	71
Table F-2.	NYS Public Schools Grade 4 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	72
Table F-3.	NYS Public Schools Grade 5 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	75
Table F-4.	NYS Public Schools Grade 6 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	77
Table F-5.	NYS Public Schools Grade 7 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	79
Table F-6.	NYS Public Schools Grade 8 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	81
Table G-1.	NYS Public Schools (Without NYC) Grade 3 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	85
Table G-2.	NYS Public Schools (Without NYC) Grade 4 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	86
Table G-3.	NYS Public Schools (Without NYC) Grade 5 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	89
Table G-4.	NYS Public Schools (Without NYC) Grade 6 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	91
Table G-5.	NYS Public Schools (Without NYC) Grade 7 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	93
Table G-6.	NYS Public Schools (Without NYC) Grade 8 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	95
Table H-1.	NYC Public Schools Grades 3–8 Mathematics Operational Test 2009: Percentages of Score Differences [Local Score Minus Audit Score]	99

I. Purpose and Scope of Audit

Purpose

The New York State Grades 3–8 Mathematics assessments consist of both multiple-choice (MC) and constructed-response (CR) items. The multiple-choice items are scored at Regional Information Centers across the state. Constructed-response items are scored by teachers at the regional scoring centers, in their districts, or in their schools. To ensure that teachers apply the same rigorous scoring standards as intended by the New York State Education Department (NYSED) and to provide evidence of inter-rater reliability, NYSED conducts annual scoring audits that involve independent rescoring of five percent of all test booklets after each test administration. This audit is conducted on a stratified random sample of schools, selected from each of the grade levels.

To help teachers in the scoring process, NYSED distributes training materials, sample student booklets for various score points, and scoring rubrics. School districts provide in-service training to teachers through the use of scoring DVDs and scoring guides provided by NYSED. Combined with this training, teachers score student booklets for each score point using scoring rubrics for the constructed-response items.

Schools identified for the 2009 audit were instructed to send their students' assessments to Pearson for rescoring. Pearson is a professional scoring company known throughout the country for their quality scoring in large-scale state assessment programs. After Pearson completed the scoring, various statistical comparisons were made to evaluate the effectiveness and accuracy of the teacher scoring process. This report contains the results from those analyses.

Scope

The Grades 3–8 mathematics assessments were administered in March 2009 throughout the state. The operational data for these assessments were collected by NYSED and included both MC and CR scores. The Regional Information Centers scored the MC items and New York State teachers scored the CR items. In June 2009, Pearson conducted the audit study by rescoring the CR items from approximately five percent of all test booklets. Pearson identified a stratified sample of schools from across the state for each of the grade levels, which equaled approximately 15,000 student test booklets. The 15,000 student assessments represented a 20% over-sampling, with the intention of attaining a minimum of 12,500 student assessments in each sample for rescoring and data analyses. A total of 83,731 mathematics test booklets were collected from sample schools and rescored in June 2009.

Audit notification letters were sent to the sample schools in March 2009 and the selected schools sent their student test booklets to Pearson for audit. Pearson rescored the CR questions and matched the audit scores with the local scores collected by NYSED. This process produced two sets of test scores for each student assessment. One set came from the local scoring performed by the New York State teachers and the second set came from the audit scoring performed by Pearson. The data analysis performed in this study consists of various comparisons between the local scores and the audit scores.

II. Selection of School Sample and Test Booklets

Audit Samples

To achieve the target audit sample of 12,500 test booklets per grade level, approximately 15,000 test booklets were sampled. Six stratified random samples of schools were selected, one for each grade, from all New York State schools with Grades 3–8 enrollment to yield the target number of test booklets. Each school was selected for audit at only one grade level. All selected schools were requested to send Pearson their mathematics test booklets for the grade level selected for audit.

Each audit sample was stratified by Need/Resource Capacity Category that consists of 7 categories. The Need/Resource Capacity Index, a measure of a district's ability to meet the needs of its students with local resources, is the ratio of the estimated poverty percentage to the combined wealth ratio. The Need/Resource Capacity (N/RC) Index divides districts into four categories: those with the highest need relative to resource capacity (High N/RC), those with average need relative to resource capacity (Average N/RC), those with less than average need relative to resource capacity (Low N/RC), and charter schools. The High N/RC districts are further subdivided into four groups (see Table 1 for definition).

Table 1. Need/Resource Capacity Category Definitions

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

The first step in the sampling procedure was to calculate the state n-counts within the seven N/R groups used for sampling. Based on school enrollment data provided by NYSED, the total number of students, by grade, was calculated for each need/resource category. Table 2 identifies the n-counts for each N/RC group by grade.

Table 2. State N-counts

State N-counts							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
New York City	63402	63314	62415	62783	65573	65539	383026
Large Cities	8249	7803	7709	7489	8249	8425	47924
High Need Urban/Suburban	16089	15895	15358	15627	16239	16389	95597
High Need Rural	11524	11432	11368	11893	12639	13278	72134
Average Need	58717	59728	60099	62181	65556	66811	373092
Low Need	29532	30436	30819	31771	32078	32129	186765
Charter	2973	2386	3182	2884	2325	1424	15174
Total	190486	190994	190950	194628	202659	203995	1173712

Once the total n-counts were calculated by code for each grade level, the proportions represented by these n-counts were calculated within each cell. The following table contains those proportions.

Table 3. Target Proportions

Target Proportions							
Need/Resource Capacity Index Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
New York City	0.33	0.33	0.33	0.32	0.32	0.32	0.33
Large Cities	0.04	0.04	0.04	0.04	0.04	0.04	0.04
High Need Urban/Suburban	0.08	0.08	0.08	0.08	0.08	0.08	0.08
High Need Rural	0.06	0.06	0.06	0.06	0.06	0.07	0.06
Average Need	0.31	0.31	0.31	0.32	0.32	0.33	0.32
Low Need	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Charter	0.02	0.01	0.02	0.01	0.01	0.01	0.01

Finally, the number of students in each cell, as determined by the target proportions, was computed. These numbers are the product of the proportions in Table 3 and 15,000, which was the target sample size. This target sample size includes a 20% over-sampling to ensure a minimum sample of 12,500. The following table summarizes these n-counts.

Table 4. Target N-counts

Target N-counts per Sample							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
New York City	4993	4972	4903	4839	4853	4819	29379
Large Cities	650	613	606	577	611	620	3675
High Need Urban/Suburban	1267	1248	1206	1204	1202	1205	7333
High Need Rural	907	898	893	917	935	976	5527
Average Need	4624	4691	4721	4792	4852	4913	28593
Low Need	2326	2390	2421	2449	2374	2362	14322
Charter	234	187	250	222	172	105	1171
Totals	15000	15000	15000	15000	15000	15000	90000

Stratified Sampling Design at the School Level

Based on the target n-counts in Table 4, schools were randomly selected by grade within each N/RC group until the desired n-count was reached. Once a school was selected for a grade level, it was removed from the selection process. This process helped ensure that a school would not be audited at more than one grade level. Some school replacements were necessary so that target n-counts were met. Table 5 lists the resulting n-counts from the school sampling.

Table 5. Selected N-counts

Selected N-counts per Sample							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
New York City	5002	4973	4909	4846	4857	4823	29410
Large Cities	646	606	599	573	631	619	3674
High Need Urban/Suburban	1275	1248	1211	1209	1194	1197	7334
High Need Rural	908	904	894	912	941	975	5534
Average Need	4623	4693	4725	4792	4851	4905	28589
Low Need	2339	2386	2422	2446	2372	2365	14330
Charter	237	185	249	239	185	118	1213
Totals	15030	14995	15009	15017	15031	15002	90084

Table 6 shows the proportions within each cell based on the selected schools. A comparison between the proportions in Table 6 with the state proportions presented in Table 3 shows a very close match, thus demonstrating that the samples at each grade level are representative of New York State's student population.

Table 6. Sample Proportions

Selected Sample Proportions							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
New York City	0.33	0.33	0.33	0.32	0.32	0.32	0.33
Large Cities	0.04	0.04	0.04	0.04	0.04	0.04	0.04
High Need Urban/Suburban	0.08	0.08	0.08	0.08	0.08	0.08	0.08
High Need Rural	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Average Need	0.31	0.31	0.31	0.32	0.32	0.33	0.32
Low Need	0.16	0.16	0.16	0.16	0.16	0.16	0.16
Charter	0.02	0.01	0.02	0.02	0.01	0.01	0.01

The schools identified in the above sampling scheme were contacted by Pearson and their test booklets were used in the audit study.

III. Data Collection and School Participation

Pearson notified 796 schools and of those, 750 schools returned materials. This represents a participation rate of 94%.

After the test booklets were scored by Pearson, the audit score file was combined with the local score file. Table 7 shows the actual n-counts in the final data files after all scoring and matching of data. Table 8 shows the actual proportions in the final data files after all scoring and matching of data.

Table 7. Obtained N-counts for Mathematics

Obtained N-counts							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Totals
New York City	5019	5011	4946	4690	4874	4347	28887
Large Cities	567	525	566	522	540	577	3297
High Need Urban/Suburban	1163	1248	946	1574	1215	692	6838
High Need Rural	910	910	869	811	692	730	4922
Average Need	4441	4402	4624	4073	4351	4503	26394
Low Need	1868	1782	1967	2409	2135	2122	12283
Charter	191	188	145	193	192	91	1000
Totals	14159	14066	14063	14272	13999	13062	83621

Table 8. Obtained Proportions for Mathematics

Obtained Proportions						
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
New York City	0.35	0.36	0.35	0.33	0.35	0.33
Large Cities	0.04	0.04	0.04	0.04	0.04	0.04
High Need Urban/Suburban	0.08	0.09	0.07	0.11	0.09	0.05
High Need Rural	0.06	0.06	0.06	0.06	0.05	0.06
Average Need	0.31	0.31	0.33	0.29	0.31	0.34
Low Need	0.13	0.13	0.14	0.17	0.15	0.16
Charter	0.01	0.01	0.01	0.01	0.01	0.01

A comparison between these proportions and the desired proportions in Table 3 shows that the data files used in each grade level closely match the intended demographics and were representative of the state. The largest difference in percents was approximately 3%.

IV. Selection and Training of Auditors

Description of How the Auditors Were Selected

Scoring directors who led the audit were content experts with degrees in the subject area or a related area. Scoring directors were also chosen based on their experience in scoring the subject area. Prior to auditor training, scoring directors reviewed the training materials provided by NYSED. Scoring directors also reviewed the FAQs listed on the NYSED website and viewed NYSED-provided DVDs containing original training presentations.

Scoring supervisors for the audit also had college degrees in the subject area or a related area. Supervisors had experience in scoring the subject area and demonstrated strong organizational abilities and communication skills.

Auditors possessed, at a minimum, a four-year college degree. They were selected to work on mathematics based on their educational qualifications and their work or scoring experience.

The high quality of the auditors and high rate of return for auditors was due in part to the scoring sites' proximity to major universities and scoring sites' access to a large pool of college graduates.

Training of Auditors

Supervisor training took place in Austin from June 8 to June 10, 2009. Supervisors trained on all booklets and all grades for which they would score. One hundred eleven auditors began training on June 11, 2009. Auditors trained on items in a single booklet, completed scoring all booklets, and then trained on a new booklet for the next grade level.

Pearson staff used only those training materials supplied by NYSED and which were used in the original scorer training. Scoring directors began training by reviewing and discussing the scoring guides for items in a booklet. Scoring directors then gave auditors the practice set(s) and auditors assigned scores to these sample responses. After auditors completed the set, scoring directors reviewed and explained expert scores for the practice booklets. Subsequent practice sets for a booklet were trained in the same manner. If auditor performance or discussion of the practice sets indicated a need for reviewing or retraining, it occurred at that time.

After discussion of the practice booklets and any necessary review, auditors completed the consistency assurance set (CAS) for that booklet. A review and discussion of the scores occurred after auditors had assigned scores to all booklets in the set. The scores achieved on the CAS determined if a trainee understood and could apply the scoring criteria. To qualify to remain on the project, a trainee had to demonstrate accuracy and consistency in scoring the CAS booklets. Trainees who were unable to demonstrate accuracy and consistency in scoring were not allowed to participate in the project.

Quality Control Procedures

Scorers were expected to meet quality standards during training and scoring. Scorers who failed to meet those quality standards were released from the project. Quality control steps taken during the project included:

- **Backreading (read behinds)** was one of the primary responsibilities of scoring directors and scoring supervisors and began immediately. Backreading is a process in which supervisors check the scores of auditors immediately after they score a booklet. It was an immediate source of information on scoring accuracy and quickly alerted scoring directors and supervisors to misconceptions at the team level, indicating the need to review or retrain. Backreading continued throughout the scoring of the project. Supervisors increased backreading focus on auditors whose scoring accuracy, based on statistical reports or backreading records, was falling below expectations.
- **Second Scoring** began immediately with 10% of responses in the audit receiving an independent score by a different auditor than the original. Second score papers were randomly generated by the system. By having a different auditor score the paper a second time, without knowledge of the score given by the original auditor, it generated the inter-rater reliability statistics to verify the accuracy of the score.
- **Reports** were available throughout the project and were monitored daily by the program manager and scoring directors. These reports included the inter-rater reliability and frequency distribution for individual auditors and for teams. Auditors whose statistics were not meeting quality expectations received retraining and had to demonstrate the ability to meet expectations in order to remain on the project.

V. Audit Procedures

Description of the Audit Procedures

In Austin, one group of auditors scored Grades 3, 5, 6, and 7. For Grades 4 and 8, the auditors were divided into two groups; the first group scored Booklet 2 only and the second group scored Booklet 3 only.

Auditors recorded their scores onto scoring monitors. Scoring monitors are scannable tracking sheets onto which auditors grid the appropriate score for the booklet. Completed scoring monitors were then scanned at regular intervals throughout the day. After monitors were scanned, reports were generated for scoring directors to review and take appropriate action based on the reports (e.g., identifying auditors with low quality statistics, identifying retraining needs).

In total, sixty-eight mathematics constructed-response items were rescored by the Pearson auditors.

VI. Data Analysis

For every test booklet used in the data analysis, there were two sets of scores. The first set of scores consisted of the multiple-choice and the constructed-response scores provided by the local scoring. The second set of scores consisted of the same multiple-choice scores and the audit scores for the constructed-response items. All data analysis and comparisons were based on these two sets of scores for each test booklet.

Inter-rater reliability requires various statistics to evaluate. A single number never provides a complete picture of the reliability. Instead one needs to examine inter-rater reliability from different aspects. To achieve that goal, several analyses were performed. Item means were calculated to provide a measure of the average agreement between the local and audit scores. An intra-class correlation was computed between the local and audit scores which provides an estimate of the scoring reliability. A weighted Kappa statistic was computed to quantify the level of agreement between the categorical data provided by the local and audit scoring. Inter-rater agreement was evaluated by examining the consensus between the local and audit scores using percent of agreement. Finally, the correlation between the total scores resulting from the local and audit scoring is computed. This provides an overall evaluation of the scoring reliability.

Item Means

The average score for each constructed-response item was computed based on the local scoring and the audit scoring. Differences between the two scores were also computed. Item means for the multiple-choice items were not examined because the same item responses were used for both the local scoring and the audit scoring.

Intra-Class Correlation

The mean intra-class correlation was computed for each item. This correlation estimates the reliability of the scoring based on an average of the local and audit scores.

Weighted Kappa

The weighted Kappa (Cohen, 1968) was calculated for each item based on the local and audit scoring. This statistic produces an estimate of the reliability of the score classifications. Weighted Kappa is a measure of quantifying levels of agreement for categorical data, item scores in the case of this study. When raters tend to assign some scores more frequently than others, the agreement rates are affected. By using the weighted Kappa, larger differences between raters are given smaller weights; therefore, this statistic can differ from the inter-rater agreement measure for certain items. In this study, lower scores were more frequently assigned than the higher scores; therefore, this statistic was evaluated only as one of the many pieces of evidence supporting the reliability of the state and school scores.

Inter-Rater Agreement

For each CR item, the difference between the local score and the audit score was computed and tallied. The total of the CR items was computed and the difference between the local scoring and audit scoring results were computed. The number of times the various differences occurred was counted and the proportions were calculated.

Two total scores were computed for each test booklet using the local scoring and audit scoring results. The correlation between these scores was also computed.

Total Score Correlation

For both the local and audit scoring results, a total score on the complete assessment was computed. Then the correlation between these total scores was computed. This statistic provides an overall measure of how scoring reliability impacted total score correlations. The amount of shared variance for the total scores when the constructed-response items were scored using the local and audit scoring methods was obtained by squaring the correlation.

VII. Results

Item Means

The average score and standard deviation for each constructed-response item was computed based on local and audit scoring results. The findings from this analysis, presented in Table 9, show a very close agreement between the local and audit scores on the mathematics CR items. Specifically, 66% of the items have exactly the same mean raw scores and 29% of the items have an absolute mean difference of 0.1.

Percent of Agreement

Table 9 contains the percent of agreement and percent of approximate agreement. Percent of approximate agreement pertains to scores where the local and audit scores differed by only one score point.

When interpreting these statistics it is important to note the impact of the maximum points possible for a given item. That is, it is more likely that the two sets of scores will have exact agreement if there are only 2 maximum points, versus an item with 3 maximum points. The total percent of agreement is the sum of the exact agreement and the approximate agreement, i.e., ratings that differ by one point. This statistic is greatly influenced by the maximum points possible. Taken collectively, the percent of exact agreement ranged from 67.0-98.5%; the total percent of agreement ranges from 96.9% to a high of 100%. Consistent with the information in the item means, the percent of agreement shows a reasonably high level of agreement between the local and audit scores.

Intra-Class Correlations

The intra-class correlation (ICC) assesses rating reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The mean intra-class correlation estimates the scoring reliability based on an average of the local and audit scores.

Generally, correlations greater than 0.60 are considered strong because they explain more than one-third of the variance. Table 9 shows that all of the items had correlations greater than 0.80. Furthermore, 82% of the items had correlations equal to or greater than 0.90. The intra-class correlations ranged from 0.80 to 0.99, showing a high degree of consistency between the local and audit scores.

Table 9. NYS Public Schools Mathematics Operational Test 2009: Inter-Rater Agreement

Grade	Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
					Exact	Approx.	Total	Local	Audit	Local	Audit		
3	26	Overall	2	14049	95.7	4.1	99.8	1.4	1.4	0.86	0.87	0.98	0.94
	27	Overall	2	14049	97.3	2.6	99.8	1.6	1.6	0.73	0.73	0.98	0.96
	28	Overall	2	14049	90.3	9.6	99.9	1.7	1.8	0.52	0.50	0.89	0.75
	29	Overall	2	14049	89.2	10.6	99.8	1.5	1.6	0.75	0.72	0.94	0.84
	30	Overall	3	14049	90.2	9.0	99.3	2.4	2.4	0.97	0.95	0.97	0.88
	31	Overall	3	14049	87.5	12.0	99.5	2.6	2.7	0.60	0.57	0.88	0.74
4	31	Overall	2	13990	89.5	10.3	99.7	1.4	1.3	0.83	0.84	0.96	0.87
	32	Overall	2	13990	98.5	1.4	99.9	1.7	1.7	0.61	0.61	0.99	0.97
	33	Overall	2	13990	93.5	6.2	99.7	1.5	1.5	0.79	0.81	0.97	0.91
	34	Overall	2	13990	91.7	7.9	99.6	1.6	1.6	0.74	0.72	0.95	0.86
	35	Overall	2	13990	88.9	10.9	99.8	1.2	1.2	0.85	0.85	0.96	0.88
	36	Overall	2	13990	93.4	6.5	99.9	1.4	1.4	0.71	0.71	0.97	0.91
	37	Overall	2	13990	91.6	8.3	99.9	1.4	1.4	0.74	0.73	0.96	0.89
	38	Overall	3	13990	72.6	26.8	99.4	1.9	1.8	1.03	1.03	0.93	0.75
	39	Overall	3	13990	93.9	5.9	99.9	2.0	2.0	0.79	0.78	0.97	0.92
	40	Overall	2	13990	97.7	2.1	99.9	1.7	1.7	0.64	0.64	0.98	0.95
	41	Overall	2	13990	88.9	11.0	99.8	1.5	1.4	0.72	0.78	0.95	0.85
	42	Overall	2	13990	96.2	3.6	99.8	1.6	1.6	0.74	0.74	0.98	0.94
	43	Overall	2	13990	94.4	5.5	99.9	1.7	1.7	0.53	0.55	0.94	0.87
	44	Overall	2	13990	93.7	6.1	99.8	1.2	1.2	0.89	0.90	0.98	0.93
	45	Overall	2	13990	95.0	4.9	99.9	1.6	1.6	0.68	0.69	0.97	0.91
	46	Overall	2	13990	92.9	7.0	99.9	1.7	1.7	0.49	0.51	0.92	0.83
	47	Overall	3	13990	90.8	8.8	99.5	2.5	2.5	0.89	0.91	0.97	0.88
	48	Overall	3	13990	94.6	5.2	99.8	2.2	2.2	0.94	0.95	0.98	0.94

Table 9. NYS Public Schools Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Grade	Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa	
					Exact	Approx.	Total	Local	Audit	Local	Audit			
5	27	Overall	2	14007	90.4	9.3	99.7	1.4	1.4	0.85	0.81	0.96	0.88	
	28	Overall	2	14007	93.0	6.8	99.8	1.5	1.5	0.70	0.71	0.96	0.90	
	29	Overall	2	14007	95.0	4.7	99.7	1.5	1.5	0.73	0.74	0.97	0.93	
	30	Overall	2	14007	92.4	6.9	99.3	1.8	1.8	0.56	0.57	0.92	0.80	
	31	Overall	3	14007	83.6	15.0	98.6	1.9	1.8	0.94	1.00	0.94	0.83	
	32	Overall	3	14007	68.5	29.9	98.4	1.5	1.4	0.83	0.83	0.85	0.63	
	33	Overall	3	14007	69.5	27.4	96.9	1.8	1.7	1.09	1.04	0.91	0.72	
	34	Overall	3	14007	75.6	22.8	98.4	2.3	2.2	1.04	1.15	0.93	0.76	
6	26	Overall	2	14203	88.0	11.3	99.3	1.7	1.7	0.63	0.64	0.90	0.75	
		27	Overall	2	14203	73.3	25.9	99.2	1.1	1.0	0.70	0.61	0.80	0.60
		28	Overall	2	14203	90.8	8.7	99.5	1.1	1.1	0.94	0.96	0.97	0.90
		29	Overall	2	14203	89.4	10.4	99.8	1.5	1.5	0.75	0.74	0.95	0.85
		30	Overall	2	14203	87.3	12.4	99.7	1.4	1.4	0.78	0.75	0.94	0.83
		31	Overall	2	14203	91.1	8.6	99.7	1.3	1.3	0.91	0.93	0.97	0.90
		32	Overall	3	14203	79.8	19.4	99.1	1.8	1.8	1.02	1.00	0.94	0.81
		33	Overall	3	14203	90.5	9.0	99.5	1.5	1.5	1.28	1.28	0.98	0.93
	34	Overall	3	14203	76.5	22.9	99.3	2.2	2.0	0.68	0.63	0.84	0.63	
	35	Overall	3	14203	94.4	5.3	99.7	2.6	2.6	0.82	0.81	0.97	0.91	
7	31	Overall	2	13885	91.9	7.5	99.4	1.7	1.7	0.68	0.69	0.94	0.84	
		32	Overall	2	13885	86.3	13.4	99.7	1.0	1.0	0.83	0.84	0.95	0.85
		33	Overall	2	13885	88.2	11.6	99.7	1.5	1.4	0.74	0.77	0.94	0.84
		34	Overall	2	13885	73.7	25.7	99.4	0.9	0.9	0.83	0.86	0.89	0.71
		35	Overall	3	13885	84.5	14.6	99.2	2.1	2.0	1.16	1.16	0.96	0.86

Table 9. NYS Public Schools Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Grade	Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
					Exact	Approx.	Total	Local	Audit	Local	Audit		
7	36	Overall	3	13885	76.7	21.5	98.2	1.2	1.3	1.23	1.25	0.95	0.81
	37	Overall	3	13885	86.2	12.7	98.9	2.0	2.0	1.27	1.29	0.97	0.88
	38	Overall	3	13885	73.4	25.9	99.3	2.3	2.1	1.03	1.01	0.93	0.73
8	28	Overall	2	12973	96.2	3.6	99.8	1.7	1.7	0.65	0.65	0.97	0.92
	29	Overall	2	12973	67.0	32.4	99.4	1.5	1.3	0.72	0.71	0.83	0.56
	30	Overall	2	12973	95.3	4.5	99.8	1.1	1.1	0.71	0.71	0.97	0.94
	31	Overall	2	12973	88.9	10.8	99.7	1.2	1.2	0.86	0.87	0.96	0.87
	32	Overall	3	12973	86.4	12.1	98.5	1.9	1.9	1.30	1.33	0.97	0.89
	33	Overall	3	12973	73.1	25.1	98.2	2.2	2.1	1.02	1.04	0.92	0.73
	34	Overall	2	12973	82.0	17.7	99.7	1.6	1.6	0.67	0.67	0.88	0.70
	35	Overall	2	12973	85.8	13.4	99.2	1.4	1.5	0.71	0.67	0.91	0.78
	36	Overall	2	12973	88.6	11.1	99.7	1.5	1.5	0.78	0.77	0.95	0.84
	37	Overall	2	12973	78.6	20.0	98.6	1.3	1.4	0.77	0.71	0.87	0.70
	38	Overall	2	12973	82.6	16.8	99.4	1.0	1.1	0.88	0.86	0.93	0.81
	39	Overall	2	12973	76.1	23.1	99.3	1.2	1.2	0.82	0.81	0.89	0.72
	40	Overall	2	12973	77.1	21.5	98.6	1.1	1.2	0.77	0.75	0.87	0.70
	41	Overall	2	12973	88.2	11.5	99.7	1.3	1.3	0.74	0.73	0.94	0.84
	42	Overall	3	12973	82.7	16.3	99.0	2.2	2.1	1.09	1.06	0.95	0.83
	43	Overall	3	12973	74.8	24.0	98.8	1.8	1.8	0.84	0.82	0.88	0.69
	44	Overall	3	12973	82.5	14.7	97.2	1.8	1.8	1.22	1.21	0.95	0.84
45	Overall	3	12973	83.5	14.8	98.3	1.8	1.8	1.31	1.32	0.97	0.87	

Approximate agreement (%) is the percent of pairs of readers that differ by one score point. Total agreement (%) is the sum of exact and approximate percents.

Weighted Kappa

The weighted Kappa is an estimate of the reliability of the score classifications. That is, the Kappa statistic is a measure of reproducibility for categorical data. A common stumbling block in evaluating scoring reliability or consistency is the basic concept of agreement beyond chance and, in turn, the importance of correcting for chance agreement. The Kappa statistic corrects for this chance agreement and tells us how much of the possible agreement over and above chance the scorers have achieved.

Guidelines for the evaluation of Kappa are:

- $k > 0.75$ denotes excellent reproducibility
- $0.4 \leq k \leq 0.75$ denotes good reproducibility
- $0 \leq k \leq 0.4$ denotes marginal reproducibility

The results found in Table 9 show a high degree of consistency between the local and audit scores. In particular, 78% of the items had a weighted Kappa statistic which denotes excellent reproducibility. The remaining 14 items produced a weighted Kappa statistic denoting good reproducibility.

Inter-Rater Agreement

For each CR item, the difference between the local score and the audit score was computed and tallied. The total of the constructed-response items was also computed and the difference between the local scoring and audit scoring totals was computed. The absolute value of the differences between the local scores and the audit scores were then tallied and the proportions computed. Those proportions are presented in Table 10. Appendices F through H contain the proportion of actual differences instead of the absolute values.

Table 10. Percentage of Absolute Raw Score Differences between Local and Audit Scores

Grade	Item	MAX Points	Difference			
			0	1	2	3
3 N=14049	26	2	96%	4%	0%	0%
	27	2	97%	3%	0%	0%
	28	2	90%	10%	0%	0%
	29	2	89%	11%	0%	0%
	30	3	90%	9%	1%	0%
	31	3	88%	12%	0%	0%
4 N=13990	31	2	89%	10%	0%	0%
	32	2	99%	1%	0%	0%
	33	2	94%	6%	0%	0%
	34	2	92%	8%	0%	0%
	35	2	89%	11%	0%	0%
	36	2	93%	7%	0%	0%
	37	2	92%	8%	0%	0%
	38	3	73%	27%	1%	0%
	39	3	94%	6%	0%	0%
	40	2	98%	2%	0%	0%
	41	2	89%	11%	0%	0%
	42	2	96%	4%	0%	0%
	43	2	94%	6%	0%	0%
	44	2	94%	6%	0%	0%
45	2	95%	5%	0%	0%	
5 N=14007	27	2	90%	9%	0%	0%
	28	2	93%	7%	0%	0%
	29	2	95%	5%	0%	0%
	30	2	92%	7%	1%	0%
	31	3	84%	15%	1%	0%
	32	3	69%	30%	2%	0%
	33	3	69%	27%	3%	0%
	34	3	76%	23%	1%	0%

Table 10. Percentage of Absolute Raw Score Differences between Local and Audit Scores (continued)

Grade	Item	MAX Points	Difference			
			0	1	2	3
6 N=14203	26	2	88%	11%	1%	0%
	27	2	73%	26%	1%	0%
	28	2	91%	9%	0%	0%
	29	2	89%	10%	0%	0%
	30	2	87%	12%	0%	0%
	31	2	91%	9%	0%	0%
	32	3	80%	19%	1%	0%
	33	3	90%	9%	0%	0%
	34	3	76%	23%	1%	0%
	35	3	94%	5%	0%	0%
7 N=13885	31	2	92%	8%	1%	0%
	32	2	86%	13%	0%	0%
	33	2	88%	12%	0%	0%
	34	2	74%	26%	1%	0%
	35	3	85%	15%	1%	0%
	36	3	77%	21%	2%	0%
	37	3	86%	13%	1%	0%
	38	3	73%	26%	1%	0%
8 N=12973	28	2	96%	4%	0%	0%
	29	2	67%	32%	1%	0%
	30	2	95%	5%	0%	0%
	31	2	89%	11%	0%	0%
	32	3	86%	12%	1%	0%
	33	3	73%	25%	2%	0%
	34	2	82%	18%	0%	0%
	35	2	86%	13%	1%	0%
	36	2	89%	11%	0%	0%
	37	2	79%	20%	1%	0%
	38	2	83%	17%	1%	0%
	39	2	76%	23%	1%	0%
	40	2	77%	21%	1%	0%
	41	2	88%	12%	0%	0%
	42	3	83%	16%	1%	0%
43	3	75%	24%	1%	0%	
44	3	83%	15%	3%	0%	
45	3	84%	15%	1%	0%	

The information provided in Table 10 shows a high degree of consistency between the local and audit scores. Specifically, the percentage of ratings that were exactly the same across local and audit scores met or exceeded 67% for all items. The percent of scores that differed by two or more points fell below 4% for all items.

Total Score Correlation

For both the local and audit scoring results, total scores on the entire assessment and the open-ended questions only were computed. Then the correlation between the local and audit total scores was computed. This statistic provides an overall measure of the scoring reliability. The amount of variance of the total scores that is shared by the local and audit scoring is obtained by squaring the correlation. This statistic is an indication of the consistency between the local scoring and audit scoring on the total-test-score level.

Table 11. Correlations Between Local and Audit Scores

Grade	Total Score		Total of Open-Ended Questions Only	
	Correlation	Common Variance	Correlation	Common Variance
3	0.99	0.98	0.97	0.94
4	0.99	0.98	0.99	0.98
5	0.99	0.98	0.96	0.92
6	0.99	0.98	0.97	0.94
7	0.99	0.98	0.97	0.94
8	0.99	0.98	0.98	0.96

The correlations show a very high degree of consistency between the local and audit scoring results with every grade having a correlation of 0.99. Based on these correlations, the amount of common variance between local and audit scores is 0.98. This means that differences in constructed-response scoring between the local and audit scoring results did not impact the total score level much. Given that most decisions using test results are based on the total score, these statistics provide valuable evidence of the reliability and consistency in students' total scores across local and audit scoring methods. The correlations based on the open-ended questions range from 0.96 to 0.99 and the common variance ranges from 0.92 to 0.98. This, again, shows a high degree of agreement.

Additional Analyses

The results from additional analyses are presented in the appendices.

Appendix A contains a detailed item analysis for the mathematics constructed-response items resulting from the local scoring. These tables show the proportion of students obtaining each of the possible score points for each item. The tables also provide the item mean and point-biserial (PBS).

The same item analysis for the mathematics audit scores are in Appendix B.

Appendices C, D, and E contain summary item-level information for the mathematics assessments. Analyses are computed for all schools and then by scoring model, where the scoring models are:

1. Regional scoring
2. Schools from two districts
3. Three or more schools within a district
4. Two schools within a district
5. Only one school

The appendices are for:

1. All schools in New York State (NYS),
2. All schools without the New York City schools, and
3. New York City (NYC) schools only.

These tables summarize the following item-level information:

- Maximum score points
- Exact agreement
- Approximate agreement
- Item mean and standard deviation from audit and local scoring
- Intra-class correlation
- Kappa statistic

Appendices F, G, and H contain the distribution of differences at the item level between the audit scores and the local scores for mathematics. This information was computed for the various scoring models. The appendices are for:

1. All schools in New York State,
2. All schools without the New York City schools, and
3. New York City schools only.

VIII. Summary

The sample acquisition was very successful. A comparison between the obtained proportions with the state proportions found in Tables 3 and 8 show that the samples mirrored the state in these categories. For all grades, the obtained proportions in each of the seven Need/Resource Capacity categories were virtually identical to the state proportions. As a result, the analysis performed in the study is based on data which is representative of the state demographics.

A summary of the analyses performed in this study indicates that the local scoring results were very close to the audit scoring results. Correlations between the total scores resulting from the audit scoring and the local scoring were 0.99 for all grade levels. The correlations based on the open-ended only range from 0.96 to 0.98. These correlations indicate a high degree of agreement between local and audit scoring results.

Examination of the differences between local scoring and audit scoring at the item level also shows a high degree of consistency. In mathematics, the largest mean difference between local and audit scoring was 0.2, which occurred in three items: Grade 6, item 34, Grade 7, item 38, and Grade 8, item 29. Considering that two of the three items are 3-point items, the difference of only 0.2 represents 7% of the maximum points for those items. All other items had an absolute mean differences of 0.1 or less.

Appendix C contains the scoring results for each of the scoring models. By examination, it appears that there is little difference between the local and audit scoring results by scoring model. Scoring models 1 and 3 had only one item where the absolute mean difference is 0.2. Scoring model 2 had 6 items where the absolute mean difference is 0.2. Scoring model 4 had 4 items where the absolute mean difference is 0.2 and scoring model 5 had 3 items. Two items had differences of 0.3 and all other items had mean differences of 0.01 or less. This shows a high degree of consistency, not only between the local and audit scoring, but also across scoring models.

In conclusion, the local scoring results are very consistent with the audit scoring. No major discrepancies were found in these analyses.

References

Cohen J. Weighted Kappa: "Nominal scale agreement with provision for scaled disagreement or partial credit." *Psychological Bulletin* 70:213–20, 1968.

Appendix A

Mathematics Item Analysis for Local Scoring

Local Scoring Mathematics Grade 3 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
26	O	0.00	0.25	0.07	0.68	0.00	0.00	1.43	0.59
27	O	0.00	0.14	0.15	0.70	0.00	0.00	1.56	0.59
28	O	0.00	0.04	0.18	0.78	0.00	0.00	1.74	0.41
29	O	0.00	0.15	0.14	0.70	0.00	0.00	1.55	0.62
30	O	0.00	0.09	0.08	0.18	0.65	0.00	2.39	0.65
31	O	0.00	0.01	0.03	0.27	0.69	0.00	2.64	0.48

Local Scoring Mathematics Grade 4 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
31	O	0.00	0.23	0.18	0.59	0.00	0.00	1.36	0.62
32	O	0.00	0.07	0.17	0.76	0.00	0.00	1.68	0.52
33	O	0.00	0.19	0.09	0.72	0.00	0.00	1.53	0.58
34	O	0.00	0.15	0.11	0.74	0.00	0.00	1.58	0.55
35	O	0.00	0.30	0.25	0.46	0.00	0.00	1.16	0.67
36	O	0.00	0.13	0.32	0.55	0.00	0.00	1.42	0.55
37	O	0.00	0.15	0.28	0.57	0.00	0.00	1.42	0.64
38	O	0.00	0.12	0.21	0.29	0.38	0.00	1.92	0.62
39	O	0.00	0.06	0.10	0.58	0.26	0.00	2.02	0.48
40	O	0.00	0.10	0.13	0.78	0.00	0.00	1.68	0.57
41	O	0.00	0.14	0.26	0.60	0.00	0.00	1.46	0.61
42	O	0.00	0.15	0.15	0.70	0.00	0.00	1.55	0.59
43	O	0.00	0.04	0.17	0.79	0.00	0.00	1.74	0.45
44	O	0.00	0.30	0.15	0.55	0.00	0.00	1.24	0.68
45	O	0.00	0.11	0.15	0.74	0.00	0.00	1.62	0.57
46	O	0.00	0.02	0.20	0.78	0.00	0.00	1.75	0.45
47	O	0.00	0.04	0.13	0.11	0.71	0.00	2.48	0.49
48	O	0.00	0.10	0.08	0.37	0.46	0.00	2.18	0.61

Local Scoring Mathematics Grade 5 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
27	O	0.00	0.24	0.13	0.63	0.00	0.00	1.39	0.62
28	O	0.00	0.12	0.27	0.60	0.00	0.00	1.48	0.57
29	O	0.00	0.14	0.24	0.62	0.00	0.00	1.48	0.63
30	O	0.00	0.07	0.10	0.83	0.00	0.00	1.77	0.38
31	O	0.00	0.12	0.15	0.45	0.28	0.00	1.89	0.55
32	O	0.00	0.10	0.43	0.36	0.11	0.00	1.48	0.62
33	O	0.00	0.17	0.20	0.28	0.35	0.00	1.80	0.62
34	O	0.00	0.08	0.19	0.09	0.64	0.00	2.29	0.58

Local Scoring Mathematics Grade 6 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
26	O	0.00	0.09	0.13	0.77	0.00	0.00	1.68	0.42
27	O	0.00	0.20	0.50	0.29	0.00	0.00	1.09	0.44
28	O	0.00	0.41	0.10	0.49	0.00	0.00	1.08	0.66
29	O	0.00	0.16	0.21	0.63	0.00	0.00	1.47	0.62
30	O	0.00	0.18	0.23	0.59	0.00	0.00	1.41	0.59
31	O	0.00	0.30	0.09	0.60	0.00	0.00	1.29	0.64
32	O	0.00	0.14	0.17	0.38	0.30	0.00	1.84	0.65
33	O	0.00	0.31	0.22	0.09	0.38	0.00	1.54	0.72
34	O	0.00	0.02	0.10	0.57	0.31	0.00	2.17	0.47
35	O	0.00	0.05	0.05	0.16	0.73	0.00	2.57	0.59

Local Scoring Mathematics Grade 7 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
31	O	0.00	0.12	0.10	0.77	0.00	0.00	1.65	0.50
32	O	0.01	0.32	0.30	0.37	0.00	0.00	1.04	0.61
33	O	0.00	0.15	0.23	0.62	0.00	0.00	1.47	0.56
34	O	0.01	0.37	0.31	0.32	0.00	0.00	0.94	0.61
35	O	0.00	0.16	0.15	0.15	0.54	0.00	2.07	0.62
36	O	0.00	0.41	0.17	0.16	0.25	0.00	1.24	0.72
37	O	0.00	0.24	0.05	0.12	0.58	0.00	2.04	0.63
38	O	0.00	0.12	0.05	0.24	0.59	0.00	2.29	0.59

Local Scoring Mathematics Grade 8 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
28	O	0.00	0.11	0.05	0.84	0.00	0.00	1.73	0.52
29	O	0.00	0.13	0.21	0.66	0.00	0.00	1.52	0.55
30	O	0.00	0.20	0.48	0.32	0.00	0.00	1.12	0.64
31	O	0.00	0.28	0.21	0.51	0.00	0.00	1.23	0.63
32	O	0.01	0.26	0.09	0.11	0.54	0.00	1.91	0.72
33	O	0.00	0.11	0.12	0.24	0.53	0.00	2.19	0.62
34	O	0.00	0.10	0.21	0.69	0.00	0.00	1.58	0.53
35	O	0.01	0.12	0.30	0.56	0.00	0.00	1.43	0.56
36	O	0.01	0.17	0.16	0.66	0.00	0.00	1.48	0.61
37	O	0.01	0.18	0.30	0.51	0.00	0.00	1.32	0.54
38	O	0.01	0.36	0.23	0.40	0.00	0.00	1.04	0.64
39	O	0.01	0.24	0.29	0.47	0.00	0.00	1.22	0.61
40	O	0.01	0.23	0.40	0.36	0.00	0.00	1.12	0.57
41	O	0.00	0.16	0.35	0.48	0.00	0.00	1.32	0.56
42	O	0.00	0.13	0.12	0.20	0.54	0.00	2.15	0.63
43	O	0.00	0.09	0.18	0.54	0.19	0.00	1.83	0.59
44	O	0.01	0.20	0.20	0.13	0.46	0.00	1.84	0.69
45	O	0.01	0.27	0.13	0.09	0.49	0.00	1.79	0.74

Appendix B

Mathematics Item Analysis for Audit Scoring

Audit Scoring Mathematics Grade 3 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
26	O	0.00	0.25	0.07	0.68	0.00	0.00	1.42	0.59
27	O	0.00	0.14	0.15	0.70	0.00	0.00	1.56	0.59
28	O	0.00	0.03	0.16	0.81	0.00	0.00	1.77	0.41
29	O	0.00	0.13	0.18	0.68	0.00	0.00	1.55	0.62
30	O	0.00	0.08	0.07	0.17	0.67	0.00	2.44	0.65
31	O	0.00	0.01	0.02	0.23	0.73	0.00	2.69	0.48

Audit Scoring Mathematics Grade 4 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
31	O	0.00	0.24	0.17	0.59	0.00	0.00	1.34	0.64
32	O	0.00	0.08	0.17	0.76	0.00	0.00	1.68	0.56
33	O	0.00	0.20	0.07	0.72	0.00	0.00	1.52	0.61
34	O	0.00	0.14	0.12	0.74	0.00	0.00	1.60	0.56
35	O	0.00	0.29	0.25	0.46	0.00	0.00	1.17	0.68
36	O	0.00	0.13	0.34	0.53	0.00	0.00	1.40	0.55
37	O	0.00	0.15	0.28	0.57	0.00	0.00	1.42	0.67
38	O	0.00	0.12	0.25	0.28	0.34	0.00	1.84	0.03
39	O	0.00	0.06	0.11	0.59	0.23	0.00	2.00	0.23
40	O	0.00	0.10	0.12	0.78	0.00	0.00	1.68	0.61
41	O	0.00	0.18	0.22	0.60	0.00	0.00	1.42	0.64
42	O	0.00	0.15	0.15	0.70	0.00	0.00	1.55	0.62
43	O	0.00	0.05	0.19	0.76	0.00	0.00	1.71	0.49
44	O	0.00	0.32	0.13	0.55	0.00	0.00	1.23	0.70
45	O	0.00	0.12	0.15	0.74	0.00	0.00	1.62	0.60
46	O	0.00	0.03	0.21	0.76	0.00	0.00	1.73	0.49
47	O	0.00	0.05	0.14	0.11	0.69	0.00	2.45	0.52
48	O	0.00	0.10	0.07	0.37	0.45	0.00	2.18	0.63

Audit Scoring Mathematics Grade 5 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
27	O	0.00	0.21	0.16	0.63	0.00	0.00	1.42	0.60
28	O	0.00	0.12	0.27	0.60	0.00	0.00	1.48	0.56
29	O	0.00	0.15	0.24	0.61	0.00	0.00	1.47	0.63
30	O	0.00	0.07	0.10	0.83	0.00	0.00	1.76	0.38
31	O	0.00	0.15	0.13	0.44	0.27	0.00	1.83	0.56
32	O	0.00	0.10	0.47	0.31	0.12	0.00	1.44	0.64
33	O	0.00	0.17	0.23	0.34	0.25	0.00	1.67	0.62
34	O	0.00	0.15	0.14	0.10	0.61	0.00	2.16	0.60

Audit Scoring Mathematics Grade 6 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
26	O	0.00	0.10	0.10	0.80	0.00	0.00	1.71	0.40
27	O	0.00	0.18	0.63	0.19	0.00	0.00	1.01	0.42
28	O	0.00	0.41	0.07	0.51	0.00	0.00	1.10	0.66
29	O	0.00	0.15	0.22	0.63	0.00	0.00	1.48	0.61
30	O	0.00	0.16	0.25	0.58	0.00	0.00	1.42	0.58
31	O	0.00	0.33	0.06	0.61	0.00	0.00	1.28	0.64
32	O	0.00	0.14	0.18	0.40	0.27	0.00	1.80	0.66
33	O	0.00	0.31	0.22	0.09	0.38	0.00	1.53	0.72
34	O	0.00	0.02	0.11	0.67	0.20	0.00	2.05	0.50
35	O	0.00	0.05	0.05	0.17	0.73	0.00	2.57	0.58

Audit Scoring Mathematics Grade 7 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
31	O	0.00	0.13	0.08	0.79	0.00	0.00	1.66	0.50
32	O	0.01	0.35	0.29	0.36	0.00	0.00	1.00	0.60
33	O	0.00	0.17	0.21	0.62	0.00	0.00	1.45	0.56
34	O	0.01	0.40	0.25	0.34	0.00	0.00	0.94	0.60
35	O	0.00	0.16	0.16	0.14	0.53	0.00	2.05	0.63
36	O	0.00	0.42	0.15	0.17	0.26	0.00	1.27	0.71
37	O	0.01	0.26	0.04	0.13	0.56	0.00	2.00	0.64
38	O	0.00	0.13	0.05	0.39	0.43	0.00	2.11	0.57

Audit Scoring Mathematics Grade 8 Item Statistics

Item	Key	B	0	1	2	3	4	Mean	Point-Biserial
28	O	0.00	0.11	0.05	0.83	0.00	0.00	1.72	0.51
29	O	0.00	0.14	0.41	0.45	0.00	0.00	1.30	0.65
30	O	0.01	0.20	0.48	0.31	0.00	0.00	1.10	0.64
31	O	0.00	0.29	0.20	0.51	0.00	0.00	1.22	0.63
32	O	0.01	0.28	0.07	0.10	0.54	0.00	1.88	0.72
33	O	0.00	0.12	0.14	0.28	0.46	0.00	2.08	0.64
34	O	0.00	0.10	0.19	0.71	0.00	0.00	1.60	0.53
35	O	0.02	0.08	0.34	0.56	0.00	0.00	1.47	0.55
36	O	0.01	0.17	0.16	0.66	0.00	0.00	1.49	0.61
37	O	0.01	0.12	0.35	0.52	0.00	0.00	1.39	0.53
38	O	0.01	0.32	0.26	0.41	0.00	0.00	1.08	0.64
39	O	0.01	0.25	0.31	0.44	0.00	0.00	1.18	0.64
40	O	0.01	0.19	0.41	0.38	0.00	0.00	1.17	0.58
41	O	0.00	0.16	0.39	0.45	0.00	0.00	1.28	0.57
42	O	0.00	0.11	0.15	0.21	0.52	0.00	2.14	0.63
43	O	0.00	0.08	0.19	0.55	0.17	0.00	1.81	0.59
44	O	0.01	0.20	0.20	0.13	0.47	0.00	1.85	0.69
45	O	0.01	0.28	0.13	0.08	0.50	0.00	1.78	0.74

Appendix C

Item Level Statistics for Mathematics Including All Schools in State

Table C-1. NYS Public Schools Grade 3 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
26	Overall	2	14049	95.7	4.1	99.8	1.4	1.4	0.86	0.87	0.98	0.94
	1	2	6799	96.1	3.6	99.8	1.4	1.4	0.89	0.89	0.99	0.95
	2	2	172	97.7	2.3	100.0	1.7	1.7	0.74	0.73	0.99	0.96
	3	2	4639	95.5	4.3	99.8	1.5	1.5	0.84	0.84	0.98	0.94
	4	2	825	95.0	4.5	99.5	1.6	1.6	0.77	0.77	0.97	0.92
	5	2	1614	94.2	5.5	99.7	1.4	1.4	0.87	0.88	0.98	0.93
27	Overall	2	14049	97.3	2.6	99.8	1.6	1.6	0.73	0.73	0.98	0.96
	1	2	6799	97.2	2.7	99.8	1.5	1.5	0.76	0.75	0.98	0.96
	2	2	172	97.7	2.3	100.0	1.8	1.8	0.61	0.60	0.98	0.95
	3	2	4639	97.7	2.2	99.9	1.6	1.6	0.70	0.70	0.99	0.96
	4	2	825	97.3	2.4	99.8	1.7	1.7	0.65	0.65	0.98	0.94
	5	2	1614	96.3	3.4	99.8	1.6	1.6	0.73	0.72	0.98	0.94
28	Overall	2	14049	90.3	9.6	99.9	1.7	1.8	0.52	0.50	0.89	0.75
	1	2	6799	90.3	9.6	99.9	1.7	1.8	0.53	0.51	0.90	0.76
	2	2	172	91.9	8.1	100.0	1.9	1.8	0.34	0.38	0.82	0.67
	3	2	4639	89.9	10.1	100.0	1.8	1.8	0.50	0.49	0.89	0.74
	4	2	825	89.3	10.7	100.0	1.8	1.8	0.47	0.44	0.85	0.69
	5	2	1614	91.8	8.1	99.8	1.8	1.8	0.51	0.50	0.90	0.78
29	Overall	2	14049	89.2	10.6	99.8	1.5	1.6	0.75	0.72	0.94	0.84
	1	2	6799	89.6	10.3	99.8	1.5	1.5	0.78	0.74	0.95	0.85
	2	2	172	90.7	9.3	100.0	1.7	1.7	0.68	0.63	0.94	0.83
	3	2	4639	89.0	10.8	99.8	1.6	1.6	0.72	0.70	0.94	0.83
	4	2	825	88.7	11.0	99.8	1.7	1.7	0.67	0.64	0.92	0.79
	5	2	1614	88.1	11.8	99.9	1.6	1.6	0.71	0.69	0.93	0.81

Table C-1. NYS Public Schools Grade 3 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
30	Overall	3	14049	90.2	9.0	99.3	2.4	2.4	0.97	0.95	0.97	0.88
	1	3	6799	89.7	9.6	99.3	2.3	2.4	1.00	0.98	0.97	0.88
	2	3	172	92.4	7.0	99.4	2.6	2.6	0.86	0.84	0.97	0.88
	3	3	4639	90.7	8.6	99.3	2.4	2.5	0.94	0.91	0.97	0.88
	4	3	825	91.2	8.0	99.2	2.6	2.6	0.79	0.76	0.95	0.85
	5	3	1614	90.2	8.9	99.1	2.4	2.4	1.00	0.96	0.96	0.88
31	Overall	3	14049	87.5	12.0	99.5	2.6	2.7	0.60	0.57	0.88	0.74
	1	3	6799	86.8	12.5	99.4	2.6	2.6	0.63	0.60	0.89	0.75
	2	3	172	94.2	5.8	100.0	2.9	2.9	0.37	0.37	0.88	0.76
	3	3	4639	87.7	11.8	99.5	2.7	2.7	0.57	0.54	0.87	0.72
	4	3	825	89.9	9.9	99.9	2.8	2.8	0.48	0.46	0.86	0.73
	5	3	1614	87.8	11.8	99.6	2.6	2.7	0.61	0.57	0.89	0.75

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Table C-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
31	Overall	2	13990	89.5	10.3	99.7	1.4	1.3	0.83	0.84	0.96	0.87
	1	2	6732	89.3	10.5	99.8	1.3	1.3	0.85	0.87	0.96	0.88
	2	2	46	82.6	15.2	97.8	1.5	1.3	0.74	0.88	0.90	0.76
	3	2	4359	90.2	9.6	99.8	1.4	1.4	0.80	0.81	0.96	0.87
	4	2	649	89.1	10.3	99.4	1.4	1.4	0.81	0.81	0.95	0.85
	5	2	2204	88.7	10.8	99.5	1.4	1.4	0.80	0.82	0.95	0.86
32	Overall	2	13990	98.5	1.4	99.9	1.7	1.7	0.61	0.61	0.99	0.97
	1	2	6732	98.2	1.6	99.9	1.6	1.6	0.65	0.65	0.99	0.97
	2	2	46	100.0	0.0	100.0	1.6	1.6	0.77	0.77	1.00	1.00
	3	2	4359	98.9	1.1	100.0	1.7	1.7	0.56	0.55	0.99	0.97
	4	2	649	98.5	1.4	99.8	1.7	1.7	0.56	0.57	0.98	0.96
	5	2	2204	98.7	1.3	100.0	1.7	1.7	0.56	0.56	0.99	0.97
33	Overall	2	13990	93.5	6.2	99.7	1.5	1.5	0.79	0.81	0.97	0.91
	1	2	6732	92.7	7.0	99.7	1.5	1.5	0.82	0.84	0.97	0.90
	2	2	46	100.0	0.0	100.0	1.7	1.7	0.68	0.68	1.00	1.00
	3	2	4359	95.0	4.8	99.7	1.6	1.6	0.75	0.76	0.97	0.92
	4	2	649	94.1	5.7	99.8	1.6	1.6	0.78	0.78	0.97	0.91
	5	2	2204	92.9	6.9	99.8	1.5	1.5	0.77	0.80	0.97	0.90
34	Overall	2	13990	91.7	7.9	99.6	1.6	1.6	0.74	0.72	0.95	0.86
	1	2	6732	89.8	9.8	99.6	1.5	1.5	0.79	0.76	0.95	0.85
	2	2	46	93.5	6.5	100.0	1.5	1.6	0.71	0.68	0.97	0.90
	3	2	4359	93.9	5.8	99.7	1.7	1.7	0.68	0.68	0.96	0.88
	4	2	649	93.2	6.2	99.4	1.7	1.7	0.62	0.66	0.94	0.85
	5	2	2204	92.6	7.0	99.6	1.6	1.7	0.70	0.67	0.95	0.86
35	Overall	2	13990	88.9	10.9	99.8	1.2	1.2	0.85	0.85	0.96	0.88
	1	2	6732	88.1	11.7	99.8	1.1	1.1	0.87	0.87	0.96	0.87
	2	2	46	95.7	4.3	100.0	1.6	1.6	0.71	0.64	0.98	0.93
	3	2	4359	90.1	9.7	99.8	1.3	1.3	0.82	0.81	0.96	0.88
	4	2	649	89.1	10.9	100.0	1.2	1.2	0.81	0.82	0.96	0.87
	5	2	2204	88.9	10.7	99.6	1.2	1.2	0.84	0.83	0.95	0.87

Table C-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
36	Overall	2	13990	93.4	6.5	99.9	1.4	1.4	0.71	0.71	0.97	0.91
	1	2	6732	93.7	6.2	99.9	1.4	1.4	0.73	0.72	0.97	0.91
	2	2	46	97.8	2.2	100.0	1.9	1.9	0.45	0.37	0.97	0.90
	3	2	4359	94.7	5.3	100.0	1.4	1.4	0.69	0.69	0.97	0.92
	4	2	649	92.4	7.6	100.0	1.6	1.5	0.68	0.66	0.96	0.88
	5	2	2204	90.3	9.6	99.9	1.4	1.4	0.72	0.71	0.95	0.87
37	Overall	2	13990	91.6	8.3	99.9	1.4	1.4	0.74	0.73	0.96	0.89
	1	2	6732	90.2	9.6	99.8	1.4	1.3	0.76	0.76	0.95	0.87
	2	2	46	97.8	2.2	100.0	1.7	1.7	0.66	0.62	0.99	0.96
	3	2	4359	93.2	6.8	100.0	1.5	1.5	0.71	0.70	0.96	0.90
	4	2	649	93.5	6.3	99.8	1.5	1.5	0.70	0.70	0.96	0.90
	5	2	2204	91.9	8.0	100.0	1.5	1.5	0.71	0.70	0.96	0.89
38	Overall	3	13990	72.6	26.8	99.4	1.9	1.8	1.03	1.03	0.93	0.75
	1	3	6732	71.3	28.0	99.3	1.9	1.8	1.05	1.05	0.92	0.75
	2	3	46	69.6	30.4	100.0	2.2	2.0	0.86	0.90	0.90	0.68
	3	3	4359	74.5	25.1	99.6	2.0	1.9	1.02	1.02	0.93	0.77
	4	3	649	74.3	25.3	99.5	2.0	2.0	0.98	0.96	0.92	0.75
	5	3	2204	72.3	27.0	99.3	1.9	1.8	1.02	1.00	0.92	0.74
39	Overall	3	13990	93.9	5.9	99.9	2.0	2.0	0.79	0.78	0.97	0.92
	1	3	6732	94.4	5.4	99.8	2.0	2.0	0.81	0.80	0.98	0.93
	2	3	46	91.3	8.7	100.0	2.4	2.4	0.68	0.71	0.95	0.87
	3	3	4359	94.2	5.7	99.9	2.0	2.0	0.76	0.74	0.97	0.92
	4	3	649	95.8	4.2	100.0	2.0	2.0	0.78	0.77	0.98	0.95
	5	3	2204	91.5	8.5	100.0	2.0	1.9	0.77	0.75	0.96	0.88
40	Overall	2	13990	97.7	2.1	99.9	1.7	1.7	0.64	0.64	0.98	0.95
	1	2	6732	97.4	2.4	99.8	1.6	1.6	0.69	0.69	0.98	0.95
	2	2	46	100.0	0.0	100.0	1.9	1.9	0.35	0.35	1.00	1.00
	3	2	4359	98.2	1.8	99.9	1.7	1.7	0.57	0.57	0.98	0.96
	4	2	649	98.3	1.7	100.0	1.7	1.7	0.60	0.60	0.99	0.96
	5	2	2204	97.8	2.0	99.8	1.7	1.7	0.58	0.59	0.98	0.94

Table C-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
41	Overall	2	13990	88.9	11.0	99.8	1.5	1.4	0.72	0.78	0.95	0.85
	1	2	6732	87.2	12.5	99.8	1.4	1.4	0.74	0.82	0.94	0.84
	2	2	46	95.7	4.3	100.0	1.7	1.7	0.58	0.53	0.96	0.90
	3	2	4359	90.2	9.7	100.0	1.5	1.5	0.70	0.73	0.95	0.86
	4	2	649	90.6	9.4	100.0	1.5	1.5	0.69	0.71	0.95	0.86
	5	2	2204	90.6	9.2	99.8	1.5	1.4	0.72	0.74	0.95	0.87
42	Overall	2	13990	96.2	3.6	99.8	1.6	1.6	0.74	0.74	0.98	0.94
	1	2	6732	95.8	4.0	99.8	1.5	1.5	0.78	0.78	0.98	0.94
	2	2	46	100.0	0.0	100.0	1.8	1.8	0.54	0.54	1.00	1.00
	3	2	4359	96.9	3.0	99.8	1.6	1.6	0.69	0.68	0.98	0.94
	4	2	649	96.8	3.2	100.0	1.6	1.6	0.69	0.68	0.98	0.95
	5	2	2204	96.0	3.8	99.8	1.6	1.6	0.70	0.70	0.98	0.93
43	Overall	2	13990	94.4	5.5	99.9	1.7	1.7	0.53	0.55	0.94	0.87
	1	2	6732	93.2	6.6	99.8	1.8	1.7	0.50	0.54	0.93	0.84
	2	2	46	95.7	2.2	97.8	1.8	1.8	0.55	0.46	0.89	0.80
	3	2	4359	95.7	4.3	100.0	1.7	1.7	0.54	0.54	0.96	0.90
	4	2	649	94.9	5.1	100.0	1.8	1.7	0.51	0.54	0.95	0.88
	5	2	2204	95.1	4.7	99.8	1.7	1.7	0.57	0.59	0.96	0.90
44	Overall	2	13990	93.7	6.1	99.8	1.2	1.2	0.89	0.90	0.98	0.93
	1	2	6732	93.4	6.5	99.8	1.1	1.1	0.91	0.93	0.98	0.93
	2	2	46	97.8	2.2	100.0	1.8	1.8	0.62	0.58	0.98	0.94
	3	2	4359	94.2	5.6	99.8	1.3	1.3	0.86	0.87	0.98	0.93
	4	2	649	95.7	4.3	100.0	1.3	1.3	0.88	0.87	0.99	0.95
	5	2	2204	92.8	6.8	99.6	1.3	1.3	0.85	0.86	0.97	0.91
45	Overall	2	13990	95.0	4.9	99.9	1.6	1.6	0.68	0.69	0.97	0.91
	1	2	6732	94.9	4.9	99.9	1.6	1.6	0.69	0.70	0.97	0.92
	2	2	46	95.7	4.3	100.0	1.9	1.9	0.32	0.25	0.85	0.65
	3	2	4359	95.7	4.2	100.0	1.7	1.7	0.67	0.66	0.97	0.92
	4	2	649	93.2	6.6	99.8	1.7	1.7	0.63	0.63	0.95	0.87
	5	2	2204	94.4	5.5	99.9	1.6	1.6	0.70	0.71	0.97	0.91

Table C-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
46	Overall	2	13990	92.9	7.0	99.9	1.7	1.7	0.49	0.51	0.92	0.83
	1	2	6732	93.6	6.3	99.9	1.8	1.7	0.48	0.51	0.93	0.84
	2	2	46	95.7	4.3	100.0	2.0	2.0	0.20	0.20	0.65	0.48
	3	2	4359	93.5	6.4	99.9	1.7	1.7	0.50	0.51	0.93	0.84
	4	2	649	92.6	6.9	99.5	1.8	1.7	0.41	0.51	0.89	0.79
	5	2	2204	89.8	10.0	99.8	1.7	1.7	0.52	0.53	0.89	0.77
47	Overall	3	13990	90.8	8.8	99.5	2.5	2.5	0.89	0.91	0.97	0.88
	1	3	6732	91.4	8.1	99.5	2.4	2.4	0.91	0.95	0.97	0.89
	2	3	46	97.8	2.2	100.0	2.8	2.8	0.63	0.59	0.99	0.95
	3	3	4359	91.2	8.4	99.6	2.5	2.5	0.85	0.86	0.96	0.88
	4	3	649	86.9	12.9	99.8	2.6	2.5	0.79	0.89	0.95	0.82
	5	3	2204	89.0	10.4	99.4	2.5	2.4	0.89	0.91	0.96	0.86
48	Overall	3	13990	94.6	5.2	99.8	2.2	2.2	0.94	0.95	0.98	0.94
	1	3	6732	93.7	6.0	99.8	2.1	2.1	0.99	0.99	0.98	0.94
	2	3	46	100.0	0.0	100.0	2.5	2.5	0.83	0.83	1.00	1.00
	3	3	4359	95.9	4.0	99.9	2.3	2.3	0.87	0.88	0.99	0.95
	4	3	649	95.4	4.3	99.7	2.3	2.3	0.92	0.92	0.98	0.94
	5	3	2204	94.2	5.5	99.7	2.2	2.2	0.92	0.91	0.98	0.93

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Table C-3. NYS Public Schools Grade 5 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
27	Overall	2	14007	90.4	9.3	99.7	1.4	1.4	0.85	0.81	0.96	0.88
	1	2	6671	90.2	9.6	99.8	1.3	1.4	0.87	0.83	0.96	0.88
	2	2	271	92.3	7.7	100.0	1.6	1.6	0.75	0.67	0.96	0.88
	3	2	3473	90.6	9.2	99.8	1.4	1.5	0.84	0.80	0.96	0.88
	4	2	618	92.9	6.8	99.7	1.5	1.5	0.77	0.74	0.96	0.89
	5	2	2974	89.8	9.7	99.5	1.4	1.4	0.83	0.81	0.96	0.87
28	Overall	2	14007	93.0	6.8	99.8	1.5	1.5	0.70	0.71	0.96	0.90
	1	2	6671	92.5	7.3	99.8	1.4	1.4	0.71	0.71	0.96	0.89
	2	2	271	92.3	7.7	100.0	1.6	1.6	0.66	0.66	0.95	0.88
	3	2	3473	93.4	6.4	99.9	1.5	1.5	0.70	0.71	0.96	0.90
	4	2	618	95.5	4.4	99.8	1.6	1.6	0.66	0.65	0.97	0.92
	5	2	2974	93.1	6.8	99.9	1.5	1.5	0.70	0.70	0.96	0.90
29	Overall	2	14007	95.0	4.7	99.7	1.5	1.5	0.73	0.74	0.97	0.93
	1	2	6671	94.0	5.6	99.6	1.4	1.4	0.76	0.77	0.97	0.92
	2	2	271	95.2	4.4	99.6	1.5	1.5	0.70	0.73	0.97	0.92
	3	2	3473	96.3	3.5	99.8	1.5	1.5	0.69	0.70	0.98	0.94
	4	2	618	96.8	3.1	99.8	1.6	1.6	0.63	0.65	0.98	0.94
	5	2	2974	95.3	4.5	99.8	1.6	1.5	0.69	0.70	0.97	0.92
30	Overall	2	14007	92.4	6.9	99.3	1.8	1.8	0.56	0.57	0.92	0.80
	1	2	6671	92.3	7.0	99.3	1.7	1.7	0.57	0.59	0.92	0.81
	2	2	271	92.3	6.6	98.9	1.7	1.7	0.59	0.62	0.92	0.81
	3	2	3473	92.2	7.2	99.4	1.8	1.8	0.55	0.57	0.92	0.79
	4	2	618	95.3	4.0	99.4	1.8	1.8	0.50	0.50	0.93	0.84
	5	2	2974	92.6	6.7	99.3	1.8	1.8	0.53	0.55	0.91	0.78
31	Overall	3	14007	83.6	15.0	98.6	1.9	1.8	0.94	1.00	0.94	0.83
	1	3	6671	85.1	13.1	98.2	1.8	1.8	0.97	1.01	0.94	0.84
	2	3	271	85.6	13.7	99.3	1.9	1.8	0.95	0.98	0.95	0.85
	3	3	3473	80.4	18.4	98.8	2.0	1.9	0.87	0.97	0.93	0.79
	4	3	618	88.2	11.0	99.2	1.8	1.8	1.00	1.01	0.96	0.88
	5	3	2974	82.8	16.2	99.0	1.9	1.9	0.91	0.99	0.94	0.82

Table C-3. NYS Public Schools Grade 5 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
32	Overall	3	14007	68.5	29.9	98.4	1.5	1.4	0.83	0.83	0.85	0.63
	1	3	6671	70.2	28.1	98.3	1.4	1.4	0.80	0.81	0.84	0.63
	2	3	271	70.1	29.5	99.6	1.5	1.3	0.76	0.78	0.85	0.63
	3	3	3473	67.5	31.0	98.5	1.5	1.5	0.84	0.85	0.85	0.63
	4	3	618	72.2	27.0	99.2	1.4	1.5	0.74	0.78	0.85	0.64
	5	3	2974	65.0	33.3	98.3	1.6	1.5	0.86	0.84	0.84	0.60
33	Overall	3	14007	69.5	27.4	96.9	1.8	1.7	1.09	1.04	0.90	0.72
	1	3	6671	70.2	27.3	97.6	1.7	1.6	1.09	1.01	0.91	0.72
	2	3	271	75.3	22.5	97.8	1.9	1.7	1.02	1.00	0.92	0.76
	3	3	3473	68.6	27.6	96.3	1.9	1.8	1.08	1.05	0.90	0.70
	4	3	618	74.8	23.3	98.1	1.9	1.8	1.08	1.07	0.93	0.77
	5	3	2974	67.1	28.4	95.6	1.9	1.8	1.09	1.06	0.89	0.69
34	Overall	3	14007	75.6	22.8	98.4	2.3	2.2	1.04	1.15	0.93	0.76
	1	3	6671	74.8	23.8	98.5	2.3	2.2	1.05	1.15	0.93	0.75
	2	3	271	73.8	24.7	98.5	2.3	2.1	1.01	1.16	0.93	0.75
	3	3	3473	78.4	20.0	98.4	2.3	2.2	1.04	1.15	0.94	0.78
	4	3	618	75.9	23.3	99.2	2.2	2.1	1.04	1.18	0.94	0.78
	5	3	2974	74.2	23.5	97.6	2.3	2.1	1.02	1.16	0.92	0.74

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Table C-4. NYS Public Schools Grade 6 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
26	Overall	2	14203	88.0	11.3	99.3	1.7	1.7	0.63	0.64	0.90	0.75
	1	2	7193	88.0	11.5	99.4	1.7	1.7	0.63	0.62	0.90	0.74
	2	2	142	93.7	6.3	100.0	1.8	1.8	0.56	0.52	0.94	0.84
	3	2	907	88.3	10.6	98.9	1.7	1.7	0.66	0.64	0.90	0.76
	4	2	861	85.9	13.0	99.0	1.7	1.7	0.66	0.67	0.89	0.72
	5	2	5100	88.3	11.0	99.3	1.7	1.7	0.63	0.65	0.91	0.76
27	Overall	2	14203	73.3	25.9	99.2	1.1	1.0	0.70	0.61	0.80	0.60
	1	2	7193	75.4	23.9	99.4	1.1	1.0	0.69	0.61	0.82	0.63
	2	2	142	69.7	28.9	98.6	1.2	1.2	0.75	0.61	0.78	0.57
	3	2	907	76.8	22.5	99.3	1.0	0.9	0.68	0.59	0.82	0.63
	4	2	861	67.0	31.5	98.5	1.2	1.1	0.74	0.60	0.74	0.52
	5	2	5100	70.7	28.3	99.1	1.0	1.0	0.71	0.60	0.77	0.56
28	Overall	2	14203	90.8	8.7	99.5	1.1	1.1	0.94	0.96	0.97	0.90
	1	2	7193	90.8	8.6	99.4	1.0	1.1	0.94	0.96	0.97	0.90
	2	2	142	97.2	2.8	100.0	1.0	1.0	0.99	0.99	0.99	0.97
	3	2	907	91.8	7.6	99.4	1.1	1.2	0.96	0.95	0.97	0.91
	4	2	861	88.5	11.3	99.8	1.2	1.2	0.90	0.93	0.96	0.87
	5	2	5100	90.8	8.9	99.6	1.1	1.1	0.94	0.96	0.97	0.90
29	Overall	2	14203	89.4	10.4	99.8	1.5	1.5	0.75	0.74	0.95	0.85
	1	2	7193	90.0	9.8	99.8	1.5	1.5	0.74	0.74	0.95	0.86
	2	2	142	95.8	3.5	99.3	1.7	1.6	0.62	0.66	0.96	0.91
	3	2	907	84.6	15.2	99.8	1.4	1.5	0.80	0.77	0.93	0.80
	4	2	861	93.5	6.5	100.0	1.6	1.6	0.63	0.68	0.96	0.89
	5	2	5100	88.7	11.2	99.9	1.4	1.5	0.78	0.76	0.95	0.85
30	Overall	2	14203	87.3	12.4	99.7	1.4	1.4	0.78	0.75	0.94	0.83
	1	2	7193	87.3	12.5	99.8	1.4	1.4	0.78	0.76	0.94	0.84
	2	2	142	88.0	12.0	100.0	1.5	1.6	0.74	0.67	0.94	0.82
	3	2	907	88.1	11.8	99.9	1.5	1.5	0.75	0.72	0.94	0.83
	4	2	861	87.9	12.0	99.9	1.5	1.6	0.72	0.68	0.93	0.81
	5	2	5100	87.1	12.5	99.6	1.4	1.5	0.78	0.75	0.94	0.82

Table C-4. NYS Public Schools Grade 6 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
31	Overall	2	14203	91.1	8.6	99.7	1.3	1.3	0.91	0.93	0.97	0.90
	1	2	7193	91.8	8.0	99.8	1.3	1.3	0.91	0.93	0.97	0.91
	2	2	142	82.4	17.6	100.0	1.5	1.4	0.73	0.88	0.93	0.77
	3	2	907	92.0	7.8	99.8	1.3	1.3	0.93	0.93	0.97	0.91
	4	2	861	92.0	7.8	99.8	1.4	1.3	0.89	0.91	0.97	0.91
	5	2	5100	90.2	9.5	99.6	1.3	1.3	0.90	0.93	0.97	0.89
32	Overall	3	14203	79.8	19.4	99.1	1.8	1.8	1.02	1.00	0.94	0.81
	1	3	7193	80.8	18.3	99.2	1.8	1.8	1.00	0.99	0.94	0.81
	2	3	142	64.8	33.1	97.9	2.0	1.8	1.00	0.89	0.87	0.63
	3	3	907	82.2	17.2	99.4	1.7	1.7	1.03	1.02	0.95	0.84
	4	3	861	74.4	24.6	99.1	2.1	1.9	0.98	0.94	0.92	0.75
	5	3	5100	79.1	19.9	99.0	1.8	1.7	1.04	1.01	0.94	0.81
33	Overall	3	14203	90.5	9.0	99.5	1.5	1.5	1.28	1.28	0.98	0.93
	1	3	7193	90.4	9.0	99.5	1.5	1.5	1.28	1.29	0.98	0.93
	2	3	142	90.8	7.7	98.6	1.5	1.5	1.23	1.24	0.98	0.92
	3	3	907	92.2	7.6	99.8	1.7	1.7	1.30	1.30	0.99	0.94
	4	3	861	90.4	9.4	99.8	1.9	1.8	1.25	1.27	0.98	0.93
	5	3	5100	90.3	9.1	99.5	1.5	1.4	1.26	1.26	0.98	0.92
34	Overall	3	14203	76.5	22.9	99.3	2.2	2.0	0.68	0.63	0.83	0.63
	1	3	7193	78.4	20.8	99.2	2.1	2.0	0.66	0.61	0.83	0.63
	2	3	142	76.1	22.5	98.6	2.3	2.2	0.64	0.59	0.77	0.59
	3	3	907	78.6	20.4	99.0	2.2	2.1	0.71	0.68	0.86	0.68
	4	3	861	76.5	23.1	99.7	2.2	2.1	0.69	0.64	0.84	0.65
	5	3	5100	73.4	26.2	99.5	2.2	2.1	0.70	0.64	0.82	0.60

Table C-4. NYS Public Schools Grade 6 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
35	Overall	3	14203	94.4	5.3	99.7	2.6	2.6	0.82	0.81	0.97	0.91
	1	3	7193	94.3	5.2	99.6	2.6	2.6	0.83	0.82	0.97	0.91
	2	3	142	96.5	3.5	100.0	2.7	2.7	0.69	0.68	0.98	0.93
	3	3	907	94.7	4.7	99.4	2.6	2.6	0.84	0.83	0.97	0.91
	4	3	861	94.5	5.5	100.0	2.6	2.6	0.81	0.76	0.98	0.91
	5	3	5100	94.5	5.4	99.8	2.6	2.6	0.82	0.81	0.98	0.91

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Table C-5. NYS Public Schools Grade 7 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
31	Overall	2	13885	91.9	7.5	99.4	1.7	1.7	0.68	0.69	0.94	0.84
	1	2	6592	91.4	7.9	99.3	1.6	1.6	0.75	0.76	0.95	0.86
	2	2	110	87.3	10.9	98.2	1.8	1.8	0.55	0.51	0.81	0.60
	3	2	872	92.0	7.8	99.8	1.6	1.6	0.70	0.69	0.95	0.86
	4	2	1332	91.1	8.2	99.2	1.7	1.7	0.65	0.67	0.93	0.81
	5	2	4979	92.7	6.7	99.5	1.8	1.8	0.57	0.57	0.93	0.80
32	Overall	2	13885	86.3	13.4	99.7	1.0	1.0	0.83	0.84	0.94	0.85
	1	2	6592	85.6	14.1	99.7	1.0	0.9	0.82	0.82	0.94	0.84
	2	2	110	90.0	10.0	100.0	1.4	1.3	0.79	0.83	0.96	0.88
	3	2	872	87.0	12.5	99.5	1.2	1.2	0.84	0.87	0.95	0.85
	4	2	1332	84.8	14.7	99.5	1.0	1.0	0.84	0.87	0.94	0.83
	5	2	4979	87.3	12.4	99.7	1.1	1.1	0.84	0.86	0.95	0.86
33	Overall	2	13885	88.2	11.6	99.7	1.5	1.4	0.74	0.77	0.94	0.84
	1	2	6592	88.0	11.8	99.7	1.4	1.4	0.77	0.79	0.94	0.84
	2	2	110	90.9	7.3	98.2	1.5	1.6	0.72	0.66	0.92	0.83
	3	2	872	89.7	10.1	99.8	1.5	1.5	0.75	0.77	0.95	0.86
	4	2	1332	89.6	10.1	99.7	1.5	1.5	0.75	0.75	0.95	0.85
	5	2	4979	87.7	12.1	99.8	1.5	1.5	0.69	0.75	0.93	0.82
34	Overall	2	13885	73.7	25.7	99.4	0.9	0.9	0.83	0.86	0.89	0.71
	1	2	6592	75.7	23.8	99.5	1.0	1.0	0.84	0.87	0.90	0.73
	2	2	110	74.5	23.6	98.2	0.8	0.7	0.76	0.78	0.85	0.67
	3	2	872	72.5	27.1	99.5	0.8	0.9	0.78	0.85	0.88	0.68
	4	2	1332	69.1	29.9	99.0	0.9	1.0	0.83	0.86	0.87	0.65
	5	2	4979	72.5	26.9	99.4	0.9	0.9	0.82	0.86	0.88	0.69
35	Overall	3	13885	84.5	14.6	99.2	2.1	2.0	1.16	1.16	0.96	0.86
	1	3	6592	85.7	13.4	99.1	2.0	1.9	1.20	1.20	0.97	0.88
	2	3	110	76.4	22.7	99.1	2.2	2.2	1.01	1.05	0.93	0.76
	3	3	872	87.5	12.0	99.5	2.1	2.1	1.16	1.15	0.97	0.89
	4	3	1332	84.4	15.1	99.5	2.0	2.1	1.16	1.16	0.96	0.86
	5	3	4979	82.7	16.4	99.1	2.2	2.2	1.08	1.09	0.95	0.83

Table C-5. NYS Public Schools Grade 7 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
36	Overall	3	13885	76.7	21.5	98.2	1.2	1.3	1.23	1.25	0.95	0.81
	1	3	6592	77.0	21.3	98.3	1.2	1.2	1.22	1.26	0.95	0.82
	2	3	110	64.5	28.2	92.7	1.2	1.3	1.21	1.21	0.88	0.67
	3	3	872	75.0	22.7	97.7	1.3	1.2	1.25	1.26	0.95	0.80
	4	3	1332	76.6	21.6	98.2	1.1	1.2	1.21	1.22	0.95	0.81
	5	3	4979	77.1	21.3	98.4	1.4	1.4	1.22	1.24	0.95	0.82
37	Overall	3	13885	86.2	12.7	98.9	2.0	2.0	1.27	1.29	0.97	0.88
	1	3	6592	86.6	12.4	99.0	2.1	2.0	1.26	1.27	0.97	0.89
	2	3	110	79.1	19.1	98.2	2.2	2.2	1.12	1.15	0.94	0.79
	3	3	872	89.7	10.0	99.7	2.1	2.1	1.28	1.27	0.98	0.92
	4	3	1332	86.0	12.8	98.7	2.0	2.0	1.28	1.30	0.97	0.88
	5	3	4979	85.4	13.4	98.7	2.0	2.0	1.28	1.30	0.97	0.88
38	Overall	3	13885	73.4	25.9	99.3	2.3	2.1	1.03	1.01	0.93	0.73
	1	3	6592	75.8	23.3	99.1	2.2	2.0	1.09	1.07	0.94	0.77
	2	3	110	70.0	30.0	100.0	2.4	2.2	0.98	0.93	0.91	0.69
	3	3	872	70.2	28.7	98.9	2.3	2.0	1.05	1.03	0.92	0.71
	4	3	1332	72.7	26.9	99.5	2.3	2.1	1.03	0.99	0.93	0.73
	5	3	4979	71.0	28.4	99.4	2.4	2.2	0.92	0.91	0.90	0.68

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Table C-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
28	Overall	2	12973	96.2	3.6	99.8	1.7	1.7	0.65	0.65	0.97	0.92
	1	2	6571	95.3	4.4	99.7	1.7	1.7	0.70	0.70	0.97	0.91
	2	2	187	99.5	0.5	100.0	1.8	1.8	0.59	0.58	1.00	0.99
	3	2	2352	96.6	3.4	99.9	1.7	1.7	0.63	0.64	0.98	0.92
	4	2	869	97.2	2.6	99.9	1.8	1.8	0.59	0.59	0.98	0.93
	5	2	2994	97.4	2.5	99.8	1.8	1.8	0.56	0.56	0.97	0.92
29	Overall	2	12973	67.0	32.4	99.4	1.5	1.3	0.72	0.71	0.80	0.56
	1	2	6571	65.1	34.2	99.4	1.5	1.2	0.75	0.72	0.80	0.56
	2	2	187	67.9	32.1	100.0	1.7	1.4	0.66	0.69	0.80	0.53
	3	2	2352	69.9	29.6	99.4	1.5	1.3	0.73	0.71	0.82	0.59
	4	2	869	69.6	30.3	99.9	1.6	1.3	0.71	0.70	0.82	0.58
	5	2	2994	68.2	31.1	99.4	1.6	1.4	0.64	0.67	0.76	0.51
30	Overall	2	12973	95.3	4.5	99.8	1.1	1.1	0.71	0.71	0.97	0.94
	1	2	6571	95.0	4.8	99.8	1.1	1.1	0.73	0.72	0.97	0.93
	2	2	187	98.4	1.6	100.0	1.3	1.3	0.68	0.69	0.99	0.98
	3	2	2352	95.5	4.4	99.9	1.2	1.2	0.71	0.72	0.98	0.94
	4	2	869	96.2	3.7	99.9	1.1	1.1	0.66	0.66	0.98	0.94
	5	2	2994	95.5	4.4	99.9	1.2	1.2	0.69	0.69	0.97	0.94
31	Overall	2	12973	88.9	10.8	99.7	1.2	1.2	0.86	0.87	0.96	0.87
	1	2	6571	88.0	11.6	99.6	1.1	1.1	0.88	0.88	0.96	0.87
	2	2	187	94.1	5.9	100.0	1.4	1.4	0.84	0.85	0.98	0.93
	3	2	2352	89.9	9.9	99.8	1.3	1.3	0.84	0.85	0.96	0.88
	4	2	869	89.9	9.8	99.7	1.3	1.3	0.85	0.85	0.96	0.88
	5	2	2994	89.5	10.3	99.8	1.4	1.4	0.80	0.82	0.96	0.87
32	Overall	3	12973	86.4	12.1	98.5	1.9	1.9	1.30	1.33	0.97	0.89
	1	3	6571	85.9	12.6	98.4	1.7	1.7	1.34	1.36	0.97	0.89
	2	3	187	89.8	10.2	100.0	2.2	2.2	1.26	1.25	0.98	0.91
	3	3	2352	87.1	11.7	98.8	2.1	2.0	1.24	1.27	0.97	0.89
	4	3	869	87.1	11.0	98.2	2.0	2.0	1.28	1.31	0.97	0.88
	5	3	2994	86.5	11.8	98.4	2.1	2.1	1.20	1.26	0.97	0.87

Table C-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
33	Overall	3	12973	73.1	25.1	98.2	2.2	2.1	1.02	1.04	0.92	0.73
	1	3	6571	72.5	25.3	97.8	2.1	2.0	1.06	1.06	0.92	0.74
	2	3	187	74.3	24.6	98.9	2.4	2.3	0.92	0.90	0.91	0.70
	3	3	2352	76.6	22.4	98.9	2.1	2.0	1.04	1.07	0.93	0.78
	4	3	869	73.9	23.9	97.8	2.5	2.3	0.87	0.95	0.89	0.67
	5	3	2994	71.2	27.2	98.4	2.4	2.3	0.92	0.94	0.89	0.67
34	Overall	2	12973	82.0	17.7	99.7	1.6	1.6	0.67	0.67	0.88	0.70
	1	2	6571	80.3	19.4	99.7	1.5	1.5	0.71	0.70	0.88	0.70
	2	2	187	81.8	18.2	100.0	1.6	1.6	0.66	0.67	0.89	0.69
	3	2	2352	82.7	17.0	99.8	1.6	1.6	0.67	0.66	0.89	0.70
	4	2	869	86.9	12.1	99.0	1.7	1.7	0.63	0.64	0.89	0.73
	5	2	2994	83.8	16.2	99.9	1.7	1.7	0.60	0.60	0.87	0.67
35	Overall	2	12973	85.8	13.4	99.2	1.4	1.5	0.71	0.67	0.90	0.78
	1	2	6571	84.6	14.5	99.1	1.4	1.4	0.74	0.68	0.90	0.78
	2	2	187	87.2	11.8	98.9	1.6	1.6	0.72	0.64	0.91	0.77
	3	2	2352	85.5	13.6	99.1	1.5	1.5	0.68	0.65	0.89	0.77
	4	2	869	86.8	12.4	99.2	1.4	1.4	0.68	0.65	0.90	0.79
	5	2	2994	88.1	11.3	99.4	1.6	1.5	0.66	0.66	0.91	0.80
36	Overall	2	12973	88.6	11.1	99.7	1.5	1.5	0.78	0.77	0.95	0.84
	1	2	6571	87.5	12.0	99.5	1.4	1.4	0.83	0.81	0.95	0.84
	2	2	187	89.3	10.7	100.0	1.5	1.5	0.73	0.72	0.95	0.84
	3	2	2352	89.5	10.2	99.7	1.6	1.6	0.73	0.74	0.94	0.84
	4	2	869	88.5	11.3	99.8	1.6	1.6	0.66	0.69	0.93	0.80
	5	2	2994	90.2	9.7	99.8	1.6	1.6	0.71	0.73	0.95	0.85
37	Overall	2	12973	78.6	20.0	98.6	1.3	1.4	0.77	0.71	0.87	0.70
	1	2	6571	77.8	20.5	98.4	1.2	1.3	0.79	0.73	0.87	0.71
	2	2	187	82.4	16.0	98.4	1.4	1.4	0.71	0.74	0.88	0.74
	3	2	2352	76.9	21.4	98.3	1.3	1.4	0.79	0.70	0.85	0.68
	4	2	869	80.1	18.5	98.6	1.4	1.5	0.71	0.68	0.86	0.70
	5	2	2994	80.9	18.4	99.3	1.4	1.5	0.72	0.67	0.87	0.71

Table C-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
38	Overall	2	12973	82.6	16.8	99.4	1.0	1.1	0.88	0.86	0.93	0.81
	1	2	6571	81.7	17.7	99.4	1.0	1.0	0.88	0.86	0.93	0.80
	2	2	187	86.6	12.8	99.5	1.2	1.3	0.90	0.86	0.95	0.85
	3	2	2352	80.5	18.5	99.0	1.1	1.1	0.88	0.86	0.92	0.78
	4	2	869	80.4	18.3	98.7	1.1	1.2	0.84	0.83	0.91	0.77
	5	2	2994	86.4	13.3	99.7	1.1	1.1	0.86	0.86	0.95	0.85
39	Overall	2	12973	76.1	23.1	99.3	1.2	1.2	0.82	0.81	0.89	0.72
	1	2	6571	76.2	23.0	99.3	1.1	1.1	0.83	0.82	0.90	0.73
	2	2	187	80.7	19.3	100.0	1.3	1.4	0.76	0.79	0.91	0.76
	3	2	2352	75.6	23.4	99.0	1.2	1.2	0.82	0.81	0.89	0.71
	4	2	869	81.6	17.4	99.0	1.3	1.3	0.81	0.81	0.91	0.77
	5	2	2994	74.4	25.1	99.6	1.4	1.3	0.75	0.77	0.87	0.67
40	Overall	2	12973	77.1	21.5	98.6	1.1	1.2	0.77	0.75	0.87	0.70
	1	2	6571	75.2	23.3	98.5	1.0	1.1	0.78	0.77	0.86	0.69
	2	2	187	87.7	10.7	98.4	1.3	1.4	0.70	0.68	0.90	0.81
	3	2	2352	75.5	23.1	98.6	1.1	1.2	0.76	0.73	0.85	0.68
	4	2	869	78.0	20.3	98.3	1.3	1.2	0.75	0.76	0.87	0.71
	5	2	2994	81.8	17.3	99.1	1.3	1.3	0.70	0.70	0.88	0.74
41	Overall	2	12973	88.2	11.5	99.7	1.3	1.3	0.74	0.73	0.94	0.84
	1	2	6571	91.6	8.1	99.7	1.2	1.2	0.74	0.74	0.96	0.89
	2	2	187	94.1	5.9	100.0	1.4	1.4	0.67	0.67	0.97	0.91
	3	2	2352	84.4	15.2	99.6	1.4	1.3	0.76	0.74	0.92	0.80
	4	2	869	90.1	9.6	99.7	1.4	1.4	0.74	0.73	0.95	0.86
	5	2	2994	82.7	17.1	99.7	1.4	1.4	0.69	0.68	0.89	0.75
42	Overall	3	12973	82.7	16.3	99.0	2.2	2.1	1.09	1.06	0.95	0.83
	1	3	6571	82.8	16.1	98.9	2.0	2.0	1.13	1.11	0.96	0.84
	2	3	187	84.0	14.4	98.4	2.3	2.4	0.98	0.93	0.94	0.81
	3	3	2352	83.2	15.9	99.2	2.1	2.1	1.11	1.09	0.96	0.85
	4	3	869	81.5	17.3	98.7	2.3	2.3	0.97	0.96	0.93	0.78
	5	3	2994	82.1	17.0	99.1	2.4	2.4	0.94	0.90	0.94	0.78

Table C-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
43	Overall	3	12973	74.8	24.0	98.8	1.8	1.8	0.84	0.82	0.88	0.69
	1	3	6571	75.0	23.8	98.8	1.7	1.7	0.89	0.86	0.90	0.72
	2	3	187	76.5	21.4	97.9	2.0	2.0	0.86	0.89	0.89	0.72
	3	3	2352	76.3	22.6	98.9	1.8	1.8	0.78	0.77	0.87	0.68
	4	3	869	74.3	24.2	98.5	1.9	1.9	0.76	0.77	0.85	0.65
	5	3	2994	73.2	25.6	98.8	2.0	1.9	0.77	0.76	0.85	0.64
44	Overall	3	12973	82.5	14.7	97.2	1.8	1.8	1.22	1.21	0.95	0.84
	1	3	6571	82.8	14.7	97.5	1.7	1.7	1.25	1.24	0.96	0.85
	2	3	187	84.0	12.3	96.3	1.9	1.9	1.22	1.23	0.95	0.85
	3	3	2352	82.4	14.4	96.7	1.9	2.0	1.19	1.18	0.95	0.83
	4	3	869	84.2	12.5	96.8	2.0	2.0	1.17	1.16	0.94	0.84
	5	3	2994	81.6	15.7	97.2	2.0	2.0	1.16	1.17	0.95	0.83
45	Overall	3	12973	83.5	14.8	98.3	1.8	1.8	1.31	1.32	0.97	0.87
	1	3	6571	83.1	15.2	98.3	1.6	1.6	1.33	1.33	0.97	0.87
	2	3	187	90.9	8.6	99.5	2.0	2.0	1.28	1.33	0.98	0.93
	3	3	2352	83.4	14.9	98.3	1.9	1.9	1.27	1.28	0.96	0.86
	4	3	869	84.8	13.2	98.0	1.9	1.9	1.28	1.32	0.96	0.87
	5	3	2994	83.6	14.6	98.2	2.0	2.0	1.24	1.30	0.96	0.86

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Appendix D

Item Level Statistics for Mathematics Including All Schools in State

Without New York City Schools

**Table D-1. NYS Public Schools (Without NYC) Grade 3 Mathematics Operational Test 2009:
Inter-Rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
26	Overall	2	8899	95.4	4.4	99.8	1.5	1.5	0.84	0.85	0.98	0.94
	1	2	1649	96.2	3.5	99.7	1.4	1.4	0.87	0.87	0.98	0.95
	2	2	172	97.7	2.3	100.0	1.7	1.7	0.74	0.73	0.99	0.96
	3	2	4639	95.5	4.3	99.8	1.5	1.5	0.84	0.84	0.98	0.94
	4	2	825	95.0	4.5	99.5	1.6	1.6	0.77	0.77	0.97	0.92
	5	2	1614	94.2	5.5	99.7	1.4	1.4	0.87	0.88	0.98	0.93
27	Overall	2	8899	97.4	2.5	99.9	1.6	1.6	0.71	0.71	0.98	0.96
	1	2	1649	97.5	2.4	99.9	1.6	1.6	0.74	0.74	0.99	0.96
	2	2	172	97.7	2.3	100.0	1.8	1.8	0.61	0.60	0.98	0.95
	3	2	4639	97.7	2.2	99.9	1.6	1.6	0.70	0.70	0.99	0.96
	4	2	825	97.3	2.4	99.8	1.7	1.7	0.65	0.65	0.98	0.94
	5	2	1614	96.3	3.4	99.8	1.6	1.6	0.73	0.72	0.98	0.94
28	Overall	2	8899	90.3	9.6	99.9	1.7	1.8	0.52	0.50	0.89	0.76
	1	2	1649	90.4	9.5	99.9	1.7	1.7	0.59	0.56	0.92	0.80
	2	2	172	91.9	8.1	100.0	1.9	1.8	0.34	0.38	0.82	0.67
	3	2	4639	89.9	10.1	100.0	1.8	1.8	0.50	0.49	0.89	0.74
	4	2	825	89.3	10.7	100.0	1.8	1.8	0.47	0.44	0.85	0.69
	5	2	1614	91.8	8.1	99.8	1.8	1.8	0.51	0.50	0.90	0.78
29	Overall	2	8899	89.1	10.7	99.8	1.6	1.6	0.72	0.70	0.94	0.83
	1	2	1649	90.4	9.5	99.9	1.5	1.5	0.75	0.73	0.95	0.86
	2	2	172	90.7	9.3	100.0	1.7	1.7	0.68	0.63	0.94	0.83
	3	2	4639	89.0	10.8	99.8	1.6	1.6	0.72	0.70	0.94	0.83
	4	2	825	88.7	11.0	99.8	1.7	1.7	0.67	0.64	0.92	0.79
	5	2	1614	88.1	11.8	99.9	1.6	1.6	0.71	0.69	0.93	0.81

**Table D-1. NYS Public Schools (Without NYC) Grade 3 Mathematics Operational Test 2009:
Inter-Rater Agreement (continued)**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
30	Overall	3	8899	90.6	8.6	99.2	2.4	2.5	0.96	0.93	0.97	0.88
	1	3	1649	90.2	9.1	99.3	2.3	2.4	1.05	1.01	0.97	0.89
	2	3	172	92.4	7.0	99.4	2.6	2.6	0.86	0.84	0.97	0.88
	3	3	4639	90.7	8.6	99.3	2.4	2.5	0.94	0.91	0.97	0.88
	4	3	825	91.2	8.0	99.2	2.6	2.6	0.79	0.76	0.95	0.85
	5	3	1614	90.2	8.9	99.1	2.4	2.4	1.00	0.96	0.96	0.88
31	Overall	3	8899	87.7	11.9	99.6	2.7	2.7	0.58	0.55	0.88	0.73
	1	3	1649	85.7	13.9	99.6	2.6	2.7	0.63	0.60	0.89	0.72
	2	3	172	94.2	5.8	100.0	2.9	2.9	0.37	0.37	0.88	0.76
	3	3	4639	87.7	11.8	99.5	2.7	2.7	0.57	0.54	0.87	0.72
	4	3	825	89.9	9.9	99.9	2.8	2.8	0.48	0.46	0.86	0.73
	5	3	1614	87.8	11.8	99.6	2.6	2.7	0.61	0.57	0.89	0.75

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Table D-2. NYS Public Schools (Without NYC) Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
31	Overall	2	8925	89.8	9.9	99.7	1.4	1.4	0.81	0.82	0.96	0.87
	1	2	1667	90.6	9.3	99.9	1.3	1.3	0.84	0.85	0.97	0.89
	2	2	46	82.6	15.2	97.8	1.5	1.3	0.74	0.88	0.90	0.76
	3	2	4359	90.2	9.6	99.8	1.4	1.4	0.80	0.81	0.96	0.87
	4	2	649	89.1	10.3	99.4	1.4	1.4	0.81	0.81	0.95	0.85
	5	2	2204	88.7	10.8	99.5	1.4	1.4	0.80	0.82	0.95	0.86
32	Overall	2	8925	98.8	1.2	99.9	1.7	1.7	0.56	0.57	0.99	0.97
	1	2	1667	98.7	1.1	99.9	1.7	1.7	0.59	0.60	0.99	0.97
	2	2	46	100.0	0.0	100.0	1.6	1.6	0.77	0.77	1.00	1.00
	3	2	4359	98.9	1.1	100.0	1.7	1.7	0.56	0.55	0.99	0.97
	4	2	649	98.5	1.4	99.8	1.7	1.7	0.56	0.57	0.98	0.96
	5	2	2204	98.7	1.3	100.0	1.7	1.7	0.56	0.56	0.99	0.97
33	Overall	2	8925	94.3	5.5	99.8	1.6	1.6	0.78	0.79	0.97	0.91
	1	2	1667	94.5	5.4	99.9	1.5	1.5	0.83	0.84	0.98	0.93
	2	2	46	100.0	0.0	100.0	1.7	1.7	0.68	0.68	1.00	1.00
	3	2	4359	95.0	4.8	99.7	1.6	1.6	0.75	0.76	0.97	0.92
	4	2	649	94.1	5.7	99.8	1.6	1.6	0.78	0.78	0.97	0.91
	5	2	2204	92.9	6.9	99.8	1.5	1.5	0.77	0.80	0.97	0.90
34	Overall	2	8925	93.0	6.7	99.7	1.6	1.7	0.69	0.68	0.96	0.87
	1	2	1667	91.0	8.7	99.7	1.6	1.6	0.75	0.71	0.95	0.85
	2	2	46	93.5	6.5	100.0	1.5	1.6	0.71	0.68	0.97	0.90
	3	2	4359	93.9	5.8	99.7	1.7	1.7	0.68	0.68	0.96	0.88
	4	2	649	93.2	6.2	99.4	1.7	1.7	0.62	0.66	0.94	0.85
	5	2	2204	92.6	7.0	99.6	1.6	1.7	0.70	0.67	0.95	0.86
35	Overall	2	8925	89.5	10.2	99.8	1.2	1.2	0.83	0.82	0.96	0.88
	1	2	1667	89.0	10.9	99.9	1.1	1.1	0.85	0.84	0.96	0.88
	2	2	46	95.7	4.3	100.0	1.6	1.6	0.71	0.64	0.98	0.93
	3	2	4359	90.1	9.7	99.8	1.3	1.3	0.82	0.81	0.96	0.88
	4	2	649	89.1	10.9	100.0	1.2	1.2	0.81	0.82	0.96	0.87
	5	2	2204	88.9	10.7	99.6	1.2	1.2	0.84	0.83	0.95	0.87

**Table D-2. NYS Public Schools (Without NYC) Grade 4 Mathematics Operational Test 2009:
Inter-Rater Agreement (continued)**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
36	Overall	2	8925	92.8	7.1	99.9	1.4	1.4	0.71	0.70	0.96	0.90
	1	2	1667	91.3	8.6	99.9	1.3	1.3	0.73	0.73	0.96	0.89
	2	2	46	97.8	2.2	100.0	1.9	1.9	0.45	0.37	0.97	0.90
	3	2	4359	94.7	5.3	100.0	1.4	1.4	0.69	0.69	0.97	0.92
	4	2	649	92.4	7.6	100.0	1.6	1.5	0.68	0.66	0.96	0.88
	5	2	2204	90.3	9.6	99.9	1.4	1.4	0.72	0.71	0.95	0.87
37	Overall	2	8925	92.4	7.5	99.9	1.5	1.5	0.71	0.71	0.96	0.89
	1	2	1667	90.5	9.4	99.8	1.4	1.4	0.72	0.73	0.95	0.87
	2	2	46	97.8	2.2	100.0	1.7	1.7	0.66	0.62	0.99	0.96
	3	2	4359	93.2	6.8	100.0	1.5	1.5	0.71	0.70	0.96	0.90
	4	2	649	93.5	6.3	99.8	1.5	1.5	0.70	0.70	0.96	0.90
	5	2	2204	91.9	8.0	100.0	1.5	1.5	0.71	0.70	0.96	0.89
38	Overall	3	8925	73.2	26.3	99.5	2.0	1.9	1.02	1.01	0.93	0.75
	1	3	1667	70.7	28.7	99.4	1.8	1.8	1.02	1.03	0.92	0.73
	2	3	46	69.6	30.4	100.0	2.2	2.0	0.86	0.90	0.90	0.68
	3	3	4359	74.5	25.1	99.6	2.0	1.9	1.02	1.02	0.93	0.77
	4	3	649	74.3	25.3	99.5	2.0	2.0	0.98	0.96	0.92	0.75
	5	3	2204	72.3	27.0	99.3	1.9	1.8	1.02	1.00	0.92	0.74
39	Overall	3	8925	93.6	6.3	99.9	2.0	2.0	0.76	0.75	0.97	0.91
	1	3	1667	94.1	5.8	99.8	1.9	1.9	0.76	0.76	0.97	0.92
	2	3	46	91.3	8.7	100.0	2.4	2.4	0.68	0.71	0.95	0.87
	3	3	4359	94.2	5.7	99.9	2.0	2.0	0.76	0.74	0.97	0.92
	4	3	649	95.8	4.2	100.0	2.0	2.0	0.78	0.77	0.98	0.95
	5	3	2204	91.5	8.5	100.0	2.0	1.9	0.77	0.75	0.96	0.88
40	Overall	2	8925	98.1	1.9	99.9	1.7	1.7	0.59	0.60	0.98	0.96
	1	2	1667	98.0	2.0	100.0	1.7	1.7	0.65	0.66	0.99	0.96
	2	2	46	100.0	0.0	100.0	1.9	1.9	0.35	0.35	1.00	1.00
	3	2	4359	98.2	1.8	99.9	1.7	1.7	0.57	0.57	0.98	0.96
	4	2	649	98.3	1.7	100.0	1.7	1.7	0.60	0.60	0.99	0.96
	5	2	2204	97.8	2.0	99.8	1.7	1.7	0.58	0.59	0.98	0.94

**Table D-2. NYS Public Schools (Without NYC) Grade 4 Mathematics Operational Test 2009:
Inter-Rater Agreement (continued)**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
41	Overall	2	8925	90.0	9.9	99.9	1.5	1.5	0.71	0.74	0.95	0.86
	1	2	1667	88.4	11.5	99.9	1.5	1.4	0.72	0.79	0.94	0.85
	2	2	46	95.7	4.3	100.0	1.7	1.7	0.58	0.53	0.96	0.90
	3	2	4359	90.2	9.7	100.0	1.5	1.5	0.70	0.73	0.95	0.86
	4	2	649	90.6	9.4	100.0	1.5	1.5	0.69	0.71	0.95	0.86
	5	2	2204	90.6	9.2	99.8	1.5	1.4	0.72	0.74	0.95	0.87
42	Overall	2	8925	96.4	3.5	99.9	1.6	1.6	0.70	0.70	0.98	0.94
	1	2	1667	95.4	4.5	99.9	1.5	1.5	0.75	0.74	0.98	0.93
	2	2	46	100.0	0.0	100.0	1.8	1.8	0.54	0.54	1.00	1.00
	3	2	4359	96.9	3.0	99.8	1.6	1.6	0.69	0.68	0.98	0.94
	4	2	649	96.8	3.2	100.0	1.6	1.6	0.69	0.68	0.98	0.95
	5	2	2204	96.0	3.8	99.8	1.6	1.6	0.70	0.70	0.98	0.93
43	Overall	2	8925	95.2	4.7	99.9	1.7	1.7	0.55	0.56	0.96	0.89
	1	2	1667	94.2	5.5	99.8	1.7	1.7	0.55	0.57	0.94	0.87
	2	2	46	95.7	2.2	97.8	1.8	1.8	0.55	0.46	0.89	0.80
	3	2	4359	95.7	4.3	100.0	1.7	1.7	0.54	0.54	0.96	0.90
	4	2	649	94.9	5.1	100.0	1.8	1.7	0.51	0.54	0.95	0.88
	5	2	2204	95.1	4.7	99.8	1.7	1.7	0.57	0.59	0.96	0.90
44	Overall	2	8925	93.9	5.8	99.8	1.3	1.3	0.87	0.88	0.98	0.93
	1	2	1667	93.8	5.9	99.7	1.2	1.2	0.91	0.92	0.98	0.93
	2	2	46	97.8	2.2	100.0	1.8	1.8	0.62	0.58	0.98	0.94
	3	2	4359	94.2	5.6	99.8	1.3	1.3	0.86	0.87	0.98	0.93
	4	2	649	95.7	4.3	100.0	1.3	1.3	0.88	0.87	0.99	0.95
	5	2	2204	92.8	6.8	99.6	1.3	1.3	0.85	0.86	0.97	0.91
45	Overall	2	8925	95.1	4.8	99.9	1.6	1.6	0.68	0.68	0.97	0.92
	1	2	1667	95.3	4.6	99.9	1.5	1.6	0.71	0.71	0.97	0.93
	2	2	46	95.7	4.3	100.0	1.9	1.9	0.32	0.25	0.85	0.65
	3	2	4359	95.7	4.2	100.0	1.7	1.7	0.67	0.66	0.97	0.92
	4	2	649	93.2	6.6	99.8	1.7	1.7	0.63	0.63	0.95	0.87
	5	2	2204	94.4	5.5	99.9	1.6	1.6	0.70	0.71	0.97	0.91

Table D-2. NYS Public Schools (Without NYC) Grade 4 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
46	Overall	2	8925	92.2	7.6	99.8	1.7	1.7	0.49	0.52	0.91	0.81
	1	2	1667	91.8	7.9	99.7	1.7	1.7	0.49	0.53	0.90	0.80
	2	2	46	95.7	4.3	100.0	2.0	2.0	0.20	0.20	0.65	0.48
	3	2	4359	93.5	6.4	99.9	1.7	1.7	0.50	0.51	0.93	0.84
	4	2	649	92.6	6.9	99.5	1.8	1.7	0.41	0.51	0.89	0.79
	5	2	2204	89.8	10.0	99.8	1.7	1.7	0.52	0.53	0.89	0.77
47	Overall	3	8925	90.3	9.2	99.5	2.5	2.5	0.89	0.90	0.96	0.87
	1	3	1667	90.9	8.3	99.3	2.3	2.3	0.98	1.00	0.97	0.90
	2	3	46	97.8	2.2	100.0	2.8	2.8	0.63	0.59	0.99	0.95
	3	3	4359	91.2	8.4	99.6	2.5	2.5	0.85	0.86	0.96	0.88
	4	3	649	86.9	12.9	99.8	2.6	2.5	0.79	0.89	0.95	0.82
	5	3	2204	89.0	10.4	99.4	2.5	2.4	0.89	0.91	0.96	0.86
48	Overall	3	8925	95.1	4.7	99.8	2.2	2.2	0.91	0.91	0.98	0.94
	1	3	1667	94.0	5.8	99.8	2.1	2.1	0.96	0.96	0.98	0.94
	2	3	46	100.0	0.0	100.0	2.5	2.5	0.83	0.83	1.00	1.00
	3	3	4359	95.9	4.0	99.9	2.3	2.3	0.87	0.88	0.99	0.95
	4	3	649	95.4	4.3	99.7	2.3	2.3	0.92	0.92	0.98	0.94
	5	3	2204	94.2	5.5	99.7	2.2	2.2	0.92	0.91	0.98	0.93

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

**Table D-3. NYS Public Schools (Without NYC) Grade 5 Mathematics Operational Test 2009:
Inter-Rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
27	Overall	2	8986	90.4	9.3	99.7	1.4	1.4	0.84	0.81	0.96	0.88
	1	2	1650	89.5	10.1	99.6	1.2	1.3	0.90	0.86	0.96	0.88
	2	2	271	92.3	7.7	100.0	1.6	1.6	0.75	0.67	0.96	0.88
	3	2	3473	90.6	9.2	99.8	1.4	1.5	0.84	0.80	0.96	0.88
	4	2	618	92.9	6.8	99.7	1.5	1.5	0.77	0.74	0.96	0.89
28	Overall	2	8986	93.1	6.7	99.9	1.5	1.5	0.70	0.70	0.96	0.90
	1	2	1650	91.9	7.9	99.8	1.4	1.4	0.71	0.71	0.96	0.88
	2	2	271	92.3	7.7	100.0	1.6	1.6	0.66	0.66	0.95	0.88
	3	2	3473	93.4	6.4	99.9	1.5	1.5	0.70	0.71	0.96	0.90
	4	2	618	95.5	4.4	99.8	1.6	1.6	0.66	0.65	0.97	0.92
29	Overall	2	8986	95.6	4.1	99.8	1.5	1.5	0.70	0.71	0.97	0.93
	1	2	1650	94.5	5.1	99.6	1.4	1.4	0.75	0.76	0.97	0.92
	2	2	271	95.2	4.4	99.6	1.5	1.5	0.70	0.73	0.97	0.92
	3	2	3473	96.3	3.5	99.8	1.5	1.5	0.69	0.70	0.98	0.94
	4	2	618	96.8	3.1	99.8	1.6	1.6	0.63	0.65	0.98	0.94
30	Overall	2	8986	92.4	7.0	99.4	1.8	1.8	0.55	0.57	0.92	0.79
	1	2	1650	91.5	8.0	99.5	1.7	1.7	0.58	0.61	0.92	0.81
	2	2	271	92.3	6.6	98.9	1.7	1.7	0.59	0.62	0.92	0.81
	3	2	3473	92.2	7.2	99.4	1.8	1.8	0.55	0.57	0.92	0.79
	4	2	618	95.3	4.0	99.4	1.8	1.8	0.50	0.50	0.93	0.84
31	Overall	3	8986	81.7	16.9	98.6	1.9	1.9	0.92	1.00	0.93	0.81
	1	3	1650	79.3	18.1	97.3	1.8	1.8	0.98	1.05	0.92	0.79
	2	3	271	85.6	13.7	99.3	1.9	1.8	0.95	0.98	0.95	0.85
	3	3	3473	80.4	18.4	98.8	2.0	1.9	0.87	0.97	0.93	0.79
	4	3	618	88.2	11.0	99.2	1.8	1.8	1.00	1.01	0.96	0.88
	5	3	2974	82.8	16.2	99.0	1.9	1.9	0.91	0.99	0.94	0.82

Table D-3. NYS Public Schools (Without NYC) Grade 5 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
32	Overall	3	8986	68.3	30.2	98.6	1.5	1.5	0.83	0.83	0.85	0.63
	1	3	1650	74.4	24.4	98.8	1.3	1.3	0.78	0.75	0.86	0.66
	2	3	271	70.1	29.5	99.6	1.5	1.3	0.76	0.78	0.85	0.63
	3	3	3473	67.5	31.0	98.5	1.5	1.5	0.84	0.85	0.85	0.63
	4	3	618	72.2	27.0	99.2	1.4	1.5	0.74	0.78	0.85	0.64
	5	3	2974	65.0	33.3	98.3	1.6	1.5	0.86	0.84	0.84	0.60
33	Overall	3	8986	69.5	27.1	96.5	1.9	1.7	1.08	1.05	0.90	0.71
	1	3	1650	72.4	25.6	98.1	1.7	1.6	1.07	1.01	0.92	0.75
	2	3	271	75.3	22.5	97.8	1.9	1.7	1.02	1.00	0.92	0.76
	3	3	3473	68.6	27.6	96.3	1.9	1.8	1.08	1.05	0.90	0.70
	4	3	618	74.8	23.3	98.1	1.9	1.8	1.08	1.07	0.93	0.77
	5	3	2974	67.1	28.4	95.6	1.9	1.8	1.09	1.06	0.89	0.69
34	Overall	3	8986	75.7	22.4	98.2	2.3	2.1	1.05	1.17	0.93	0.76
	1	3	1650	73.2	25.0	98.1	2.2	2.0	1.12	1.19	0.93	0.76
	2	3	271	73.8	24.7	98.5	2.3	2.1	1.01	1.16	0.93	0.75
	3	3	3473	78.4	20.0	98.4	2.3	2.2	1.04	1.15	0.94	0.78
	4	3	618	75.9	23.3	99.2	2.2	2.1	1.04	1.18	0.94	0.78
	5	3	2974	74.2	23.5	97.6	2.3	2.1	1.02	1.16	0.92	0.74

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

**Table D-4. NYS Public Schools (Without NYC) Grade 6 Mathematics Operational Test 2009:
Inter-Rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
26	Overall	2	9380	88.0	11.3	99.3	1.7	1.7	0.63	0.64	0.90	0.75
	1	2	2370	87.6	11.9	99.5	1.7	1.7	0.63	0.62	0.90	0.74
	2	2	142	93.7	6.3	100.0	1.8	1.8	0.56	0.52	0.94	0.84
	3	2	907	88.3	10.6	98.9	1.7	1.7	0.66	0.64	0.90	0.76
	4	2	861	85.9	13.0	99.0	1.7	1.7	0.66	0.67	0.89	0.72
	5	2	5100	88.3	11.0	99.3	1.7	1.7	0.63	0.65	0.91	0.76
27	Overall	2	9380	71.2	27.8	99.0	1.1	1.0	0.71	0.60	0.78	0.57
	1	2	2370	71.7	27.3	98.9	1.1	1.0	0.72	0.59	0.78	0.58
	2	2	142	69.7	28.9	98.6	1.2	1.2	0.75	0.61	0.78	0.57
	3	2	907	76.8	22.5	99.3	1.0	0.9	0.68	0.59	0.82	0.63
	4	2	861	67.0	31.5	98.5	1.2	1.1	0.74	0.60	0.74	0.52
	5	2	5100	70.7	28.3	99.1	1.0	1.0	0.71	0.60	0.77	0.56
28	Overall	2	9380	91.0	8.7	99.6	1.1	1.1	0.94	0.96	0.97	0.90
	1	2	2370	91.6	8.0	99.6	1.1	1.1	0.95	0.96	0.97	0.91
	2	2	142	97.2	2.8	100.0	1.0	1.0	0.99	0.99	0.99	0.97
	3	2	907	91.8	7.6	99.4	1.1	1.2	0.96	0.95	0.97	0.91
	4	2	861	88.5	11.3	99.8	1.2	1.2	0.90	0.93	0.96	0.87
	5	2	5100	90.8	8.9	99.6	1.1	1.1	0.94	0.96	0.97	0.90
29	Overall	2	9380	89.3	10.5	99.9	1.5	1.5	0.75	0.74	0.95	0.85
	1	2	2370	90.7	9.1	99.7	1.6	1.6	0.67	0.71	0.94	0.85
	2	2	142	95.8	3.5	99.3	1.7	1.6	0.62	0.66	0.96	0.91
	3	2	907	84.6	15.2	99.8	1.4	1.5	0.80	0.77	0.93	0.80
	4	2	861	93.5	6.5	100.0	1.6	1.6	0.63	0.68	0.96	0.89
	5	2	5100	88.7	11.2	99.9	1.4	1.5	0.78	0.76	0.95	0.85
30	Overall	2	9380	87.4	12.3	99.7	1.5	1.5	0.76	0.73	0.94	0.82
	1	2	2370	87.6	12.2	99.9	1.5	1.5	0.71	0.71	0.93	0.82
	2	2	142	88.0	12.0	100.0	1.5	1.6	0.74	0.67	0.94	0.82
	3	2	907	88.1	11.8	99.9	1.5	1.5	0.75	0.72	0.94	0.83
	4	2	861	87.9	12.0	99.9	1.5	1.6	0.72	0.68	0.93	0.81
	5	2	5100	87.1	12.5	99.6	1.4	1.5	0.78	0.75	0.94	0.82

**Table D-4. NYS Public Schools (Without NYC) Grade 6 Mathematics Operational Test 2009:
Inter-Rater Agreement (continued)**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
31	Overall	2	9380	90.6	9.1	99.7	1.3	1.3	0.90	0.93	0.97	0.89
	1	2	2370	91.1	8.7	99.7	1.3	1.3	0.91	0.92	0.97	0.90
	2	2	142	82.4	17.6	100.0	1.5	1.4	0.73	0.88	0.93	0.77
	3	2	907	92.0	7.8	99.8	1.3	1.3	0.93	0.93	0.97	0.91
	4	2	861	92.0	7.8	99.8	1.4	1.3	0.89	0.91	0.97	0.91
	5	2	5100	90.2	9.5	99.6	1.3	1.3	0.90	0.93	0.97	0.89
32	Overall	3	9380	80.1	19.1	99.2	1.8	1.8	1.02	0.99	0.94	0.81
	1	3	2370	84.2	15.4	99.6	1.9	1.9	0.96	0.96	0.95	0.84
	2	3	142	64.8	33.1	97.9	2.0	1.8	1.00	0.89	0.87	0.63
	3	3	907	82.2	17.2	99.4	1.7	1.7	1.03	1.02	0.95	0.84
	4	3	861	74.4	24.6	99.1	2.1	1.9	0.98	0.94	0.92	0.75
	5	3	5100	79.1	19.9	99.0	1.8	1.7	1.04	1.01	0.94	0.81
33	Overall	3	9380	90.8	8.7	99.5	1.5	1.5	1.27	1.28	0.98	0.93
	1	3	2370	91.5	8.1	99.6	1.6	1.6	1.29	1.28	0.98	0.94
	2	3	142	90.8	7.7	98.6	1.5	1.5	1.23	1.24	0.98	0.92
	3	3	907	92.2	7.6	99.8	1.7	1.7	1.30	1.30	0.99	0.94
	4	3	861	90.4	9.4	99.8	1.9	1.8	1.25	1.27	0.98	0.93
	5	3	5100	90.3	9.1	99.5	1.5	1.4	1.26	1.26	0.98	0.92
34	Overall	3	9380	75.7	23.8	99.5	2.2	2.1	0.69	0.64	0.83	0.63
	1	3	2370	79.2	20.3	99.5	2.1	2.1	0.65	0.64	0.84	0.66
	2	3	142	76.1	22.5	98.6	2.3	2.2	0.64	0.59	0.77	0.59
	3	3	907	78.6	20.4	99.0	2.2	2.1	0.71	0.68	0.86	0.68
	4	3	861	76.5	23.1	99.7	2.2	2.1	0.69	0.64	0.84	0.65
	5	3	5100	73.4	26.2	99.5	2.2	2.1	0.70	0.64	0.82	0.60

**Table D-4. NYS Public Schools (Without NYC) Grade 6 Mathematics Operational Test 2009:
Inter-Rater Agreement (continued)**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
35	Overall	3	9380	94.5	5.3	99.8	2.6	2.6	0.81	0.80	0.98	0.91
	1	3	2370	94.4	5.3	99.7	2.6	2.6	0.79	0.79	0.97	0.91
	2	3	142	96.5	3.5	100.0	2.7	2.7	0.69	0.68	0.98	0.93
	3	3	907	94.7	4.7	99.4	2.6	2.6	0.84	0.83	0.97	0.91
	4	3	861	94.5	5.5	100.0	2.6	2.6	0.81	0.76	0.98	0.91
	5	3	5100	94.5	5.4	99.8	2.6	2.6	0.82	0.81	0.98	0.91

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

**Table D-5. NYS Public Schools (Without NYC) Grade 7 Mathematics Operational Test 2009:
Inter-Rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
31	Overall	2	8925	93.0	6.6	99.5	1.8	1.8	0.59	0.59	0.93	0.82
	1	2	1632	96.1	3.7	99.8	1.8	1.8	0.49	0.50	0.95	0.86
	2	2	110	87.3	10.9	98.2	1.8	1.8	0.55	0.51	0.81	0.60
	3	2	872	92.0	7.8	99.8	1.6	1.6	0.70	0.69	0.95	0.86
	4	2	1332	91.1	8.2	99.2	1.7	1.7	0.65	0.67	0.93	0.81
	5	2	4979	92.7	6.7	99.5	1.8	1.8	0.57	0.57	0.93	0.80
32	Overall	2	8925	86.2	13.4	99.6	1.1	1.1	0.84	0.86	0.94	0.85
	1	2	1632	83.3	16.1	99.4	1.3	1.2	0.80	0.83	0.93	0.80
	2	2	110	90.0	10.0	100.0	1.4	1.3	0.79	0.83	0.96	0.88
	3	2	872	87.0	12.5	99.5	1.2	1.2	0.84	0.87	0.95	0.85
	4	2	1332	84.8	14.7	99.5	1.0	1.0	0.84	0.87	0.94	0.83
	5	2	4979	87.3	12.4	99.7	1.1	1.1	0.84	0.86	0.95	0.86
33	Overall	2	8925	89.2	10.5	99.8	1.5	1.5	0.69	0.73	0.94	0.84
	1	2	1632	93.3	6.7	99.9	1.7	1.7	0.54	0.62	0.95	0.86
	2	2	110	90.9	7.3	98.2	1.5	1.6	0.72	0.66	0.92	0.83
	3	2	872	89.7	10.1	99.8	1.5	1.5	0.75	0.77	0.95	0.86
	4	2	1332	89.6	10.1	99.7	1.5	1.5	0.75	0.75	0.95	0.85
	5	2	4979	87.7	12.1	99.8	1.5	1.5	0.69	0.75	0.93	0.82
34	Overall	2	8925	72.8	26.5	99.3	1.0	1.0	0.82	0.86	0.89	0.70
	1	2	1632	77.0	22.5	99.4	1.2	1.1	0.82	0.86	0.90	0.74
	2	2	110	74.5	23.6	98.2	0.8	0.7	0.76	0.78	0.85	0.67
	3	2	872	72.5	27.1	99.5	0.8	0.9	0.78	0.85	0.88	0.68
	4	2	1332	69.1	29.9	99.0	0.9	1.0	0.83	0.86	0.87	0.65
	5	2	4979	72.5	26.9	99.4	0.9	0.9	0.82	0.86	0.88	0.69
35	Overall	3	8925	83.7	15.6	99.3	2.2	2.2	1.09	1.10	0.96	0.84
	1	3	1632	84.5	15.1	99.6	2.4	2.3	0.99	1.01	0.95	0.83
	2	3	110	76.4	22.7	99.1	2.2	2.2	1.01	1.05	0.93	0.76
	3	3	872	87.5	12.0	99.5	2.1	2.1	1.16	1.15	0.97	0.89
	4	3	1332	84.4	15.1	99.5	2.0	2.1	1.16	1.16	0.96	0.86
	5	3	4979	82.7	16.4	99.1	2.2	2.2	1.08	1.09	0.95	0.83

Table D-5. NYS Public Schools (Without NYC) Grade 7 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
36	Overall	3	8925	75.5	22.6	98.2	1.3	1.4	1.23	1.25	0.95	0.81
	1	3	1632	70.9	27.1	98.0	1.4	1.6	1.22	1.26	0.94	0.77
	2	3	110	64.5	28.2	92.7	1.2	1.3	1.21	1.21	0.88	0.67
	3	3	872	75.0	22.7	97.7	1.3	1.2	1.25	1.26	0.95	0.80
	4	3	1332	76.6	21.6	98.2	1.1	1.2	1.21	1.22	0.95	0.81
	5	3	4979	77.1	21.3	98.4	1.4	1.4	1.22	1.24	0.95	0.82
37	Overall	3	8925	86.4	12.6	99.0	2.1	2.1	1.25	1.27	0.97	0.88
	1	3	1632	88.4	11.1	99.5	2.3	2.3	1.12	1.14	0.97	0.88
	2	3	110	79.1	19.1	98.2	2.2	2.2	1.12	1.15	0.94	0.79
	3	3	872	89.7	10.0	99.7	2.1	2.1	1.28	1.27	0.98	0.92
	4	3	1332	86.0	12.8	98.7	2.0	2.0	1.28	1.30	0.97	0.88
	5	3	4979	85.4	13.4	98.7	2.0	2.0	1.28	1.30	0.97	0.88
38	Overall	3	8925	71.4	28.0	99.4	2.4	2.2	0.93	0.92	0.90	0.68
	1	3	1632	72.4	27.0	99.4	2.6	2.4	0.77	0.79	0.86	0.62
	2	3	110	70.0	30.0	100.0	2.4	2.2	0.98	0.93	0.91	0.69
	3	3	872	70.2	28.7	98.9	2.3	2.0	1.05	1.03	0.92	0.71
	4	3	1332	72.7	26.9	99.5	2.3	2.1	1.03	0.99	0.93	0.73
	5	3	4979	71.0	28.4	99.4	2.4	2.2	0.92	0.91	0.90	0.68

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

**Table D-6. NYS Public Schools (Without NYC) Grade 8 Mathematics Operational Test 2009:
Inter-Rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
28	Overall	2	8656	97.3	2.6	99.9	1.8	1.8	0.57	0.58	0.98	0.92
	1	2	2254	97.8	2.1	100.0	1.8	1.8	0.52	0.52	0.98	0.93
	2	2	187	99.5	0.5	100.0	1.8	1.8	0.59	0.58	1.00	0.99
	3	2	2352	96.6	3.4	99.9	1.7	1.7	0.63	0.64	0.98	0.92
	4	2	869	97.2	2.6	99.9	1.8	1.8	0.59	0.59	0.98	0.93
	5	2	2994	97.4	2.5	99.8	1.8	1.8	0.56	0.56	0.97	0.92
29	Overall	2	8656	68.9	30.6	99.5	1.6	1.4	0.67	0.68	0.79	0.54
	1	2	2254	68.6	30.9	99.5	1.6	1.4	0.62	0.65	0.75	0.50
	2	2	187	67.9	32.1	100.0	1.7	1.4	0.66	0.69	0.80	0.53
	3	2	2352	69.9	29.6	99.4	1.5	1.3	0.73	0.71	0.82	0.59
	4	2	869	69.6	30.3	99.9	1.6	1.3	0.71	0.70	0.82	0.58
	5	2	2994	68.2	31.1	99.4	1.6	1.4	0.64	0.67	0.76	0.51
30	Overall	2	8656	95.8	4.1	99.9	1.2	1.2	0.69	0.69	0.98	0.94
	1	2	2254	96.1	3.8	99.9	1.2	1.2	0.68	0.68	0.98	0.94
	2	2	187	98.4	1.6	100.0	1.3	1.3	0.68	0.69	0.99	0.98
	3	2	2352	95.5	4.4	99.9	1.2	1.2	0.71	0.72	0.98	0.94
	4	2	869	96.2	3.7	99.9	1.1	1.1	0.66	0.66	0.98	0.94
	5	2	2994	95.5	4.4	99.9	1.2	1.2	0.69	0.69	0.97	0.94
31	Overall	2	8656	90.2	9.5	99.8	1.4	1.4	0.81	0.82	0.96	0.88
	1	2	2254	91.3	8.4	99.8	1.5	1.5	0.77	0.77	0.96	0.88
	2	2	187	94.1	5.9	100.0	1.4	1.4	0.84	0.85	0.98	0.93
	3	2	2352	89.9	9.9	99.8	1.3	1.3	0.84	0.85	0.96	0.88
	4	2	869	89.9	9.8	99.7	1.3	1.3	0.85	0.85	0.96	0.88
	5	2	2994	89.5	10.3	99.8	1.4	1.4	0.80	0.82	0.96	0.87
32	Overall	3	8656	87.8	10.8	98.6	2.2	2.1	1.20	1.24	0.97	0.88
	1	3	2254	90.3	8.7	98.9	2.4	2.3	1.07	1.14	0.97	0.89
	2	3	187	89.8	10.2	100.0	2.2	2.2	1.26	1.25	0.98	0.91
	3	3	2352	87.1	11.7	98.8	2.1	2.0	1.24	1.27	0.97	0.89
	4	3	869	87.1	11.0	98.2	2.0	2.0	1.28	1.31	0.97	0.88
	5	3	2994	86.5	11.8	98.4	2.1	2.1	1.20	1.26	0.97	0.87

Table D-6. NYS Public Schools (Without NYC) Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
33	Overall	3	8656	74.6	24.0	98.6	2.3	2.2	0.94	0.97	0.91	0.72
	1	3	2254	77.4	21.4	98.8	2.5	2.4	0.84	0.88	0.90	0.71
	2	3	187	74.3	24.6	98.9	2.4	2.3	0.92	0.90	0.91	0.70
	3	3	2352	76.6	22.4	98.9	2.1	2.0	1.04	1.07	0.93	0.78
	4	3	869	73.9	23.9	97.8	2.5	2.3	0.87	0.95	0.89	0.67
	5	3	2994	71.2	27.2	98.4	2.4	2.3	0.92	0.94	0.89	0.67
34	Overall	2	8656	82.3	17.4	99.8	1.7	1.7	0.62	0.62	0.86	0.66
	1	2	2254	78.3	21.4	99.8	1.7	1.7	0.58	0.57	0.80	0.56
	2	2	187	81.8	18.2	100.0	1.6	1.6	0.66	0.67	0.89	0.69
	3	2	2352	82.7	17.0	99.8	1.6	1.6	0.67	0.66	0.89	0.70
	4	2	869	86.9	12.1	99.0	1.7	1.7	0.63	0.64	0.89	0.73
	5	2	2994	83.8	16.2	99.9	1.7	1.7	0.60	0.60	0.87	0.67
35	Overall	2	8656	86.7	12.6	99.2	1.5	1.5	0.68	0.65	0.90	0.78
	1	2	2254	85.8	13.4	99.2	1.5	1.6	0.70	0.63	0.90	0.76
	2	2	187	87.2	11.8	98.9	1.6	1.6	0.72	0.64	0.91	0.77
	3	2	2352	85.5	13.6	99.1	1.5	1.5	0.68	0.65	0.89	0.77
	4	2	869	86.8	12.4	99.2	1.4	1.4	0.68	0.65	0.90	0.79
	5	2	2994	88.1	11.3	99.4	1.6	1.5	0.66	0.66	0.91	0.80
36	Overall	2	8656	90.0	9.7	99.7	1.6	1.6	0.69	0.71	0.94	0.83
	1	2	2254	90.8	8.7	99.5	1.7	1.7	0.62	0.62	0.92	0.80
	2	2	187	89.3	10.7	100.0	1.5	1.5	0.73	0.72	0.95	0.84
	3	2	2352	89.5	10.2	99.7	1.6	1.6	0.73	0.74	0.94	0.84
	4	2	869	88.5	11.3	99.8	1.6	1.6	0.66	0.69	0.93	0.80
	5	2	2994	90.2	9.7	99.8	1.6	1.6	0.71	0.73	0.95	0.85
37	Overall	2	8656	79.9	18.8	98.7	1.4	1.5	0.74	0.68	0.86	0.70
	1	2	2254	81.5	17.0	98.5	1.5	1.5	0.72	0.66	0.86	0.70
	2	2	187	82.4	16.0	98.4	1.4	1.4	0.71	0.74	0.88	0.74
	3	2	2352	76.9	21.4	98.3	1.3	1.4	0.79	0.70	0.85	0.68
	4	2	869	80.1	18.5	98.6	1.4	1.5	0.71	0.68	0.86	0.70
	5	2	2994	80.9	18.4	99.3	1.4	1.5	0.72	0.67	0.87	0.71

**Table D-6. NYS Public Schools (Without NYC) Grade 8 Mathematics Operational Test 2009:
Inter-Rater Agreement (continued)**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
38	Overall	2	8656	83.6	15.7	99.3	1.2	1.2	0.85	0.85	0.93	0.81
	1	2	2254	83.9	15.5	99.3	1.3	1.3	0.80	0.80	0.92	0.80
	2	2	187	86.6	12.8	99.5	1.2	1.3	0.90	0.86	0.95	0.85
	3	2	2352	80.5	18.5	99.0	1.1	1.1	0.88	0.86	0.92	0.78
	4	2	869	80.4	18.3	98.7	1.1	1.2	0.84	0.83	0.91	0.77
	5	2	2994	86.4	13.3	99.7	1.1	1.1	0.86	0.86	0.95	0.85
39	Overall	2	8656	76.6	22.8	99.4	1.3	1.3	0.78	0.79	0.89	0.71
	1	2	2254	78.1	21.5	99.6	1.3	1.3	0.77	0.77	0.89	0.73
	2	2	187	80.7	19.3	100.0	1.3	1.4	0.76	0.79	0.91	0.76
	3	2	2352	75.6	23.4	99.0	1.2	1.2	0.82	0.81	0.89	0.71
	4	2	869	81.6	17.4	99.0	1.3	1.3	0.81	0.81	0.91	0.77
	5	2	2994	74.4	25.1	99.6	1.4	1.3	0.75	0.77	0.87	0.67
40	Overall	2	8656	78.5	20.1	98.6	1.2	1.3	0.74	0.72	0.86	0.70
	1	2	2254	76.7	21.6	98.3	1.2	1.3	0.76	0.71	0.85	0.68
	2	2	187	87.7	10.7	98.4	1.3	1.4	0.70	0.68	0.90	0.81
	3	2	2352	75.5	23.1	98.6	1.1	1.2	0.76	0.73	0.85	0.68
	4	2	869	78.0	20.3	98.3	1.3	1.2	0.75	0.76	0.87	0.71
	5	2	2994	81.8	17.3	99.1	1.3	1.3	0.70	0.70	0.88	0.74
41	Overall	2	8656	85.9	13.8	99.7	1.4	1.4	0.71	0.70	0.92	0.80
	1	2	2254	89.4	10.4	99.8	1.4	1.4	0.67	0.67	0.93	0.84
	2	2	187	94.1	5.9	100.0	1.4	1.4	0.67	0.67	0.97	0.91
	3	2	2352	84.4	15.2	99.6	1.4	1.3	0.76	0.74	0.92	0.80
	4	2	869	90.1	9.6	99.7	1.4	1.4	0.74	0.73	0.95	0.86
	5	2	2994	82.7	17.1	99.7	1.4	1.4	0.69	0.68	0.89	0.75
42	Overall	3	8656	83.7	15.4	99.1	2.3	2.3	0.99	0.97	0.95	0.82
	1	3	2254	86.9	12.2	99.1	2.5	2.5	0.89	0.88	0.94	0.82
	2	3	187	84.0	14.4	98.4	2.3	2.4	0.98	0.93	0.94	0.81
	3	3	2352	83.2	15.9	99.2	2.1	2.1	1.11	1.09	0.96	0.85
	4	3	869	81.5	17.3	98.7	2.3	2.3	0.97	0.96	0.93	0.78
	5	3	2994	82.1	17.0	99.1	2.4	2.4	0.94	0.90	0.94	0.78

Table D-6. NYS Public Schools (Without NYC) Grade 8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
43	Overall	3	8656	74.9	23.9	98.8	2.0	1.9	0.76	0.77	0.86	0.66
	1	3	2254	75.8	23.2	99.0	2.0	2.0	0.72	0.74	0.85	0.65
	2	3	187	76.5	21.4	97.9	2.0	2.0	0.86	0.89	0.89	0.72
	3	3	2352	76.3	22.6	98.9	1.8	1.8	0.78	0.77	0.87	0.68
	4	3	869	74.3	24.2	98.5	1.9	1.9	0.76	0.77	0.85	0.65
	5	3	2994	73.2	25.6	98.8	2.0	1.9	0.77	0.76	0.85	0.64
44	Overall	3	8656	82.0	15.2	97.2	2.0	2.0	1.15	1.15	0.95	0.83
	1	3	2254	81.3	16.7	98.0	2.2	2.2	1.08	1.07	0.94	0.81
	2	3	187	84.0	12.3	96.3	1.9	1.9	1.22	1.23	0.95	0.85
	3	3	2352	82.4	14.4	96.7	1.9	2.0	1.19	1.18	0.95	0.83
	4	3	869	84.2	12.5	96.8	2.0	2.0	1.17	1.16	0.94	0.84
	5	3	2994	81.6	15.7	97.2	2.0	2.0	1.16	1.17	0.95	0.83
45	Overall	3	8656	84.8	13.7	98.5	2.0	2.0	1.24	1.27	0.97	0.87
	1	3	2254	87.3	11.7	99.0	2.2	2.2	1.15	1.17	0.97	0.87
	2	3	187	90.9	8.6	99.5	2.0	2.0	1.28	1.33	0.98	0.93
	3	3	2352	83.4	14.9	98.3	1.9	1.9	1.27	1.28	0.96	0.86
	4	3	869	84.8	13.2	98.0	1.9	1.9	1.28	1.32	0.96	0.87
	5	3	2994	83.6	14.6	98.2	2.0	2.0	1.24	1.30	0.96	0.86

Approximate agreement (%) is the percent of pairs of readers that differ by one score point.

Total agreement (%) is the sum of exact and approximate percents.

The scoring models are: 1) Regional scoring; 2) Schools from two districts; 3) Three or more schools within a district; 4) Two schools within a district; and 5) Only one school.

Appendix E

Item Level Statistics for Mathematics Including New York City Schools Only

Table E-1. NYC Public Schools Grades 3–8 Mathematics Operational Test 2009: Inter-Rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
Grade 3												
26	NYC	2	5150	96.1	3.7	99.8	1.4	1.4	0.89	0.89	0.99	0.95
27	NYC	2	5150	97.1	2.7	99.8	1.5	1.5	0.76	0.75	0.98	0.96
28	NYC	2	5150	90.3	9.6	99.9	1.7	1.8	0.52	0.49	0.89	0.75
29	NYC	2	5150	89.3	10.5	99.8	1.5	1.5	0.78	0.74	0.95	0.85
30	NYC	3	5150	89.6	9.7	99.3	2.3	2.4	0.99	0.97	0.96	0.88
31	NYC	3	5150	87.2	12.1	99.3	2.6	2.6	0.62	0.60	0.89	0.75
Grade 4												
31	NYC	2	5065	88.8	10.9	99.8	1.3	1.3	0.85	0.88	0.96	0.87
32	NYC	2	5065	98.0	1.8	99.8	1.6	1.6	0.67	0.67	0.99	0.96
33	NYC	2	5065	92.1	7.6	99.6	1.5	1.5	0.81	0.84	0.97	0.89
34	NYC	2	5065	89.4	10.2	99.6	1.5	1.5	0.81	0.78	0.95	0.85
35	NYC	2	5065	87.7	12.0	99.8	1.1	1.1	0.88	0.88	0.96	0.87
36	NYC	2	5065	94.4	5.4	99.9	1.4	1.4	0.72	0.72	0.97	0.92
37	NYC	2	5065	90.1	9.7	99.8	1.3	1.3	0.78	0.77	0.95	0.88
38	NYC	3	5065	71.5	27.8	99.2	1.9	1.8	1.06	1.06	0.93	0.75
39	NYC	3	5065	94.5	5.3	99.8	2.1	2.1	0.82	0.81	0.98	0.93
40	NYC	2	5065	97.2	2.5	99.7	1.6	1.6	0.71	0.70	0.98	0.95
41	NYC	2	5065	86.9	12.9	99.7	1.4	1.3	0.75	0.83	0.94	0.83
42	NYC	2	5065	95.9	3.8	99.7	1.5	1.5	0.79	0.79	0.98	0.94
43	NYC	2	5065	92.9	7.0	99.9	1.8	1.7	0.49	0.53	0.92	0.82
44	NYC	2	5065	93.2	6.6	99.9	1.1	1.1	0.91	0.93	0.98	0.93
45	NYC	2	5065	94.8	5.0	99.9	1.6	1.6	0.68	0.69	0.97	0.91
46	NYC	2	5065	94.2	5.8	99.9	1.8	1.7	0.48	0.50	0.93	0.85
47	NYC	3	5065	91.5	8.0	99.6	2.5	2.4	0.89	0.93	0.97	0.89
48	NYC	3	5065	93.7	6.1	99.7	2.1	2.1	1.00	1.01	0.98	0.94

Table E-1. NYC Public Schools Grades 3–8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
Grade 5												
27	NYC	2	5021	90.5	9.4	99.8	1.4	1.4	0.86	0.82	0.96	0.88
28	NYC	2	5021	92.7	7.1	99.8	1.4	1.4	0.71	0.72	0.96	0.90
29	NYC	2	5021	93.8	5.8	99.6	1.4	1.4	0.77	0.77	0.97	0.92
30	NYC	2	5021	92.5	6.7	99.2	1.7	1.7	0.57	0.58	0.92	0.81
31	NYC	3	5021	87.0	11.4	98.4	1.8	1.8	0.97	1.00	0.95	0.86
32	NYC	3	5021	68.8	29.3	98.1	1.4	1.4	0.81	0.83	0.84	0.62
33	NYC	3	5021	69.5	27.9	97.4	1.7	1.5	1.09	1.00	0.90	0.72
34	NYC	3	5021	75.3	23.4	98.7	2.3	2.2	1.03	1.13	0.93	0.75
Grade 6												
26	NYC	2	4823	88.1	11.2	99.4	1.7	1.7	0.63	0.62	0.90	0.74
27	NYC	2	4823	77.3	22.3	99.6	1.1	1.0	0.67	0.62	0.84	0.66
28	NYC	2	4823	90.4	8.9	99.3	1.0	1.0	0.94	0.96	0.97	0.90
29	NYC	2	4823	89.6	10.2	99.8	1.4	1.4	0.77	0.75	0.95	0.86
30	NYC	2	4823	87.2	12.6	99.8	1.3	1.3	0.81	0.79	0.94	0.84
31	NYC	2	4823	92.2	7.7	99.8	1.3	1.3	0.91	0.93	0.97	0.91
32	NYC	3	4823	79.2	19.8	99.0	1.8	1.8	1.02	1.01	0.94	0.80
33	NYC	3	4823	89.9	9.5	99.4	1.5	1.5	1.28	1.29	0.98	0.92
34	NYC	3	4823	78.0	21.0	99.1	2.1	2.0	0.67	0.60	0.82	0.62
35	NYC	3	4823	94.3	5.2	99.5	2.5	2.5	0.84	0.83	0.97	0.91
Grade 7												
31	NYC	2	4960	89.9	9.3	99.2	1.5	1.5	0.80	0.81	0.95	0.86
32	NYC	2	4960	86.4	13.4	99.8	0.9	0.9	0.81	0.80	0.94	0.84
33	NYC	2	4960	86.3	13.4	99.7	1.3	1.3	0.81	0.82	0.94	0.83
34	NYC	2	4960	75.2	24.2	99.5	0.9	0.9	0.84	0.86	0.90	0.73
35	NYC	3	4960	86.1	12.8	99.0	1.9	1.8	1.24	1.23	0.97	0.89
36	NYC	3	4960	79.0	19.4	98.3	1.1	1.1	1.21	1.24	0.95	0.83
37	NYC	3	4960	85.9	12.8	98.8	2.0	1.9	1.29	1.30	0.97	0.88
38	NYC	3	4960	76.9	22.1	99.1	2.1	1.9	1.15	1.13	0.95	0.80

Table E-1. NYC Public Schools Grades 3–8 Mathematics Operational Test 2009: Inter-Rater Agreement (continued)

Item #	Scoring Model	Score Points	Total N	Agreement (%)			Raw Score Mean		Raw Score Standard Deviation		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
Grade 8												
28	NYC	2	4317	94.0	5.6	99.6	1.6	1.6	0.76	0.76	0.97	0.90
29	NYC	2	4317	63.3	36.0	99.3	1.4	1.1	0.79	0.74	0.81	0.56
30	NYC	2	4317	94.4	5.4	99.8	1.0	1.0	0.73	0.72	0.97	0.93
31	NYC	2	4317	86.2	13.2	99.5	0.9	0.9	0.88	0.88	0.95	0.85
32	NYC	3	4317	83.6	14.6	98.1	1.4	1.4	1.33	1.35	0.97	0.87
33	NYC	3	4317	69.9	27.3	97.2	1.9	1.8	1.11	1.09	0.91	0.73
34	NYC	2	4317	81.4	18.3	99.7	1.4	1.5	0.75	0.75	0.90	0.75
35	NYC	2	4317	83.9	15.1	99.0	1.3	1.3	0.75	0.69	0.90	0.78
36	NYC	2	4317	85.8	13.8	99.6	1.2	1.3	0.87	0.85	0.95	0.84
37	NYC	2	4317	75.9	22.4	98.3	1.1	1.2	0.80	0.74	0.86	0.69
38	NYC	2	4317	80.6	18.9	99.4	0.8	0.8	0.86	0.83	0.92	0.78
39	NYC	2	4317	75.3	23.8	99.1	1.1	1.0	0.85	0.83	0.89	0.72
40	NYC	2	4317	74.4	24.2	98.6	0.9	1.0	0.78	0.77	0.86	0.68
41	NYC	2	4317	92.7	6.9	99.6	1.1	1.1	0.76	0.76	0.96	0.91
42	NYC	3	4317	80.7	18.1	98.8	1.8	1.8	1.18	1.15	0.96	0.84
43	NYC	3	4317	74.6	24.1	98.7	1.6	1.6	0.94	0.89	0.90	0.73
44	NYC	3	4317	83.6	13.6	97.2	1.5	1.5	1.27	1.25	0.96	0.86
45	NYC	3	4317	81.0	17.0	98.0	1.3	1.3	1.31	1.30	0.96	0.85

Approximate agreement (%) is the percent of pairs of readers that differ by one score point. Total agreement (%) is the sum of exact and approximate percents.

Appendix F

Item Level Differences for Mathematics Including

All Schools in State

**Table F-1. NYS Public Schools Grade 3 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
26	Overall	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	1	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	2	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	3	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	4	0.00	0.00	0.01	0.95	0.03	0.00	0.00
	5	0.00	0.00	0.02	0.94	0.03	0.00	0.00
27	Overall	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	1	0.00	0.00	0.02	0.97	0.01	0.00	0.00
	2	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	3	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	4	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	5	0.00	0.00	0.02	0.96	0.01	0.00	0.00
28	Overall	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	1	0.00	0.00	0.07	0.90	0.03	0.00	0.00
	2	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	3	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	4	0.00	0.00	0.07	0.89	0.04	0.00	0.00
	5	0.00	0.00	0.05	0.92	0.04	0.00	0.00
29	Overall	0.00	0.00	0.05	0.89	0.05	0.00	0.00
	1	0.00	0.00	0.06	0.90	0.05	0.00	0.00
	2	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	3	0.00	0.00	0.05	0.89	0.06	0.00	0.00
	4	0.00	0.00	0.06	0.89	0.05	0.00	0.00
	5	0.00	0.00	0.05	0.88	0.06	0.00	0.00
30	Overall	0.00	0.01	0.06	0.90	0.03	0.00	0.00
	1	0.00	0.01	0.07	0.90	0.03	0.00	0.00
	2	0.00	0.01	0.04	0.92	0.03	0.00	0.00
	3	0.00	0.01	0.06	0.91	0.03	0.00	0.00
	4	0.00	0.00	0.06	0.91	0.02	0.00	0.00
	5	0.00	0.01	0.06	0.90	0.03	0.00	0.00
31	Overall	0.00	0.00	0.08	0.88	0.04	0.00	0.00
	1	0.00	0.00	0.09	0.87	0.04	0.00	0.00
	2	0.00	0.00	0.03	0.94	0.02	0.00	0.00
	3	0.00	0.00	0.08	0.88	0.04	0.00	0.00
	4	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	5	0.00	0.00	0.09	0.88	0.03	0.00	0.00

**Table F-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
31	Overall	0.00	0.00	0.04	0.89	0.06	0.00	0.00
	1	0.00	0.00	0.04	0.89	0.06	0.00	0.00
	2	0.00	0.00	0.00	0.83	0.15	0.02	0.00
	3	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	4	0.00	0.00	0.06	0.89	0.05	0.01	0.00
	5	0.00	0.00	0.03	0.89	0.07	0.00	0.00
32	Overall	0.00	0.00	0.01	0.99	0.01	0.00	0.00
	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.99	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.98	0.01	0.00	0.00
	5	0.00	0.00	0.01	0.99	0.01	0.00	0.00
33	Overall	0.00	0.00	0.03	0.94	0.04	0.00	0.00
	1	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.03	0.95	0.02	0.00	0.00
	4	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.04	0.00	0.00
34	Overall	0.00	0.00	0.05	0.92	0.03	0.00	0.00
	1	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	2	0.00	0.00	0.07	0.93	0.00	0.00	0.00
	3	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.93	0.04	0.00	0.00
	5	0.00	0.00	0.05	0.93	0.02	0.00	0.00
35	Overall	0.00	0.00	0.06	0.89	0.05	0.00	0.00
	1	0.00	0.00	0.06	0.88	0.06	0.00	0.00
	2	0.00	0.00	0.04	0.96	0.00	0.00	0.00
	3	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	4	0.00	0.00	0.05	0.89	0.06	0.00	0.00
	5	0.00	0.00	0.06	0.89	0.04	0.00	0.00
36	Overall	0.00	0.00	0.02	0.93	0.04	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.92	0.05	0.00	0.00
	5	0.00	0.00	0.04	0.90	0.06	0.00	0.00

**Table F-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
37	Overall	0.00	0.00	0.04	0.92	0.04	0.00	0.00
	1	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.03	0.93	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	5	0.00	0.00	0.04	0.92	0.04	0.00	0.00
38	Overall	0.00	0.00	0.10	0.73	0.17	0.00	0.00
	1	0.00	0.00	0.10	0.71	0.18	0.00	0.00
	2	0.00	0.00	0.04	0.70	0.26	0.00	0.00
	3	0.00	0.00	0.09	0.74	0.16	0.00	0.00
	4	0.00	0.00	0.10	0.74	0.15	0.00	0.00
	5	0.00	0.00	0.09	0.72	0.18	0.00	0.00
39	Overall	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.00	0.02	0.91	0.07	0.00	0.00
	3	0.00	0.00	0.01	0.94	0.04	0.00	0.00
	4	0.00	0.00	0.02	0.96	0.03	0.00	0.00
	5	0.00	0.00	0.02	0.91	0.07	0.00	0.00
40	Overall	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	1	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	4	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	5	0.00	0.00	0.01	0.98	0.01	0.00	0.00
41	Overall	0.00	0.00	0.03	0.89	0.08	0.00	0.00
	1	0.00	0.00	0.03	0.87	0.10	0.00	0.00
	2	0.00	0.00	0.04	0.96	0.00	0.00	0.00
	3	0.00	0.00	0.04	0.90	0.06	0.00	0.00
	4	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	5	0.00	0.00	0.04	0.91	0.05	0.00	0.00
42	Overall	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	1	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.97	0.02	0.00	0.00
	4	0.00	0.00	0.02	0.97	0.01	0.00	0.00
	5	0.00	0.00	0.02	0.96	0.02	0.00	0.00

**Table F-2. NYS Public Schools Grade 4 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
43	Overall	0.00	0.00	0.01	0.94	0.04	0.00	0.00
	1	0.00	0.00	0.01	0.93	0.06	0.00	0.00
	2	0.00	0.02	0.02	0.96	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.96	0.03	0.00	0.00
	4	0.00	0.00	0.00	0.95	0.05	0.00	0.00
	5	0.00	0.00	0.01	0.95	0.03	0.00	0.00
44	Overall	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	1	0.00	0.00	0.02	0.93	0.05	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.02	0.94	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.04	0.00	0.00
45	Overall	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	2	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	3	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	4	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	5	0.00	0.00	0.02	0.94	0.03	0.00	0.00
46	Overall	0.00	0.00	0.02	0.93	0.05	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	3	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	4	0.00	0.00	0.01	0.93	0.06	0.00	0.00
	5	0.00	0.00	0.04	0.90	0.06	0.00	0.00
47	Overall	0.00	0.00	0.03	0.91	0.06	0.00	0.00
	1	0.00	0.00	0.02	0.91	0.06	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	4	0.00	0.00	0.02	0.87	0.10	0.00	0.00
	5	0.00	0.00	0.04	0.89	0.07	0.00	0.00
48	Overall	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	4	0.00	0.00	0.02	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.94	0.03	0.00	0.00

**Table F-3. NYS Public Schools Grade 5 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
27	Overall	0.00	0.00	0.06	0.90	0.03	0.00	0.00
	1	0.00	0.00	0.07	0.90	0.03	0.00	0.00
	2	0.00	0.00	0.06	0.92	0.01	0.00	0.00
	3	0.00	0.00	0.06	0.91	0.03	0.00	0.00
	4	0.00	0.00	0.04	0.93	0.03	0.00	0.00
	5	0.00	0.00	0.06	0.90	0.04	0.00	0.00
28	Overall	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	1	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	2	0.00	0.00	0.04	0.92	0.04	0.00	0.00
	3	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	4	0.00	0.00	0.03	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.03	0.00	0.00
29	Overall	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.00	0.01	0.95	0.04	0.00	0.00
	3	0.00	0.00	0.01	0.96	0.02	0.00	0.00
	4	0.00	0.00	0.01	0.97	0.02	0.00	0.00
	5	0.00	0.00	0.02	0.95	0.03	0.00	0.00
30	Overall	0.00	0.00	0.03	0.92	0.04	0.00	0.00
	1	0.00	0.00	0.03	0.92	0.04	0.00	0.00
	2	0.00	0.00	0.02	0.92	0.05	0.01	0.00
	3	0.00	0.00	0.03	0.92	0.04	0.00	0.00
	4	0.00	0.00	0.02	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.04	0.00	0.00
31	Overall	0.00	0.01	0.05	0.84	0.10	0.01	0.00
	1	0.00	0.01	0.05	0.85	0.08	0.01	0.00
	2	0.00	0.00	0.05	0.86	0.08	0.00	0.00
	3	0.00	0.00	0.04	0.80	0.14	0.01	0.00
	4	0.00	0.00	0.03	0.88	0.08	0.00	0.00
	5	0.00	0.01	0.05	0.83	0.11	0.00	0.00
32	Overall	0.00	0.01	0.13	0.69	0.17	0.01	0.00
	1	0.00	0.01	0.13	0.70	0.15	0.01	0.00
	2	0.00	0.00	0.07	0.70	0.23	0.00	0.00
	3	0.00	0.01	0.14	0.67	0.17	0.01	0.00
	4	0.00	0.00	0.18	0.72	0.09	0.00	0.00
	5	0.00	0.01	0.10	0.65	0.23	0.01	0.00

**Table F-3. NYS Public Schools Grade 5 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
33	Overall	0.00	0.01	0.09	0.69	0.18	0.02	0.00
	1	0.00	0.01	0.09	0.70	0.18	0.02	0.00
	2	0.00	0.00	0.06	0.75	0.17	0.02	0.00
	3	0.00	0.01	0.09	0.69	0.18	0.03	0.00
	4	0.00	0.01	0.09	0.75	0.14	0.01	0.00
	5	0.00	0.01	0.09	0.67	0.20	0.04	0.00
34	Overall	0.00	0.00	0.06	0.76	0.17	0.01	0.00
	1	0.00	0.00	0.06	0.75	0.17	0.01	0.00
	2	0.00	0.00	0.06	0.74	0.19	0.01	0.00
	3	0.00	0.01	0.06	0.78	0.14	0.01	0.00
	4	0.00	0.00	0.06	0.76	0.18	0.01	0.00
	5	0.00	0.01	0.05	0.74	0.18	0.01	0.00

**Table F-4. NYS Public Schools Grade 6 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
26	Overall	0.00	0.00	0.07	0.88	0.05	0.00	0.00
	1	0.00	0.00	0.07	0.88	0.04	0.00	0.00
	2	0.00	0.00	0.05	0.94	0.01	0.00	0.00
	3	0.00	0.01	0.07	0.88	0.03	0.00	0.00
	4	0.00	0.00	0.07	0.86	0.06	0.01	0.00
	5	0.00	0.00	0.06	0.88	0.05	0.00	0.00
27	Overall	0.00	0.00	0.09	0.73	0.17	0.00	0.00
	1	0.00	0.00	0.07	0.75	0.17	0.00	0.00
	2	0.00	0.01	0.10	0.70	0.19	0.00	0.00
	3	0.00	0.00	0.10	0.77	0.13	0.00	0.00
	4	0.00	0.01	0.11	0.67	0.21	0.01	0.00
	5	0.00	0.00	0.12	0.71	0.17	0.01	0.00
28	Overall	0.00	0.00	0.05	0.91	0.04	0.00	0.00
	1	0.00	0.00	0.05	0.91	0.03	0.00	0.00
	2	0.00	0.00	0.02	0.97	0.01	0.00	0.00
	3	0.00	0.00	0.05	0.92	0.02	0.00	0.00
	4	0.00	0.00	0.07	0.89	0.05	0.00	0.00
	5	0.00	0.00	0.05	0.91	0.04	0.00	0.00
29	Overall	0.00	0.00	0.06	0.89	0.05	0.00	0.00
	1	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	2	0.00	0.00	0.01	0.96	0.03	0.01	0.00
	3	0.00	0.00	0.09	0.85	0.06	0.00	0.00
	4	0.00	0.00	0.01	0.93	0.05	0.00	0.00
	5	0.00	0.00	0.07	0.89	0.04	0.00	0.00
30	Overall	0.00	0.00	0.07	0.87	0.06	0.00	0.00
	1	0.00	0.00	0.06	0.87	0.06	0.00	0.00
	2	0.00	0.00	0.11	0.88	0.01	0.00	0.00
	3	0.00	0.00	0.07	0.88	0.05	0.00	0.00
	4	0.00	0.00	0.08	0.88	0.04	0.00	0.00
	5	0.00	0.00	0.07	0.87	0.06	0.00	0.00
31	Overall	0.00	0.00	0.03	0.91	0.05	0.00	0.00
	1	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	2	0.00	0.00	0.04	0.82	0.14	0.00	0.00
	3	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	4	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	5	0.00	0.00	0.03	0.90	0.06	0.00	0.00

**Table F-4. NYS Public Schools Grade 6 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
32	Overall	0.00	0.00	0.08	0.80	0.12	0.01	0.00
	1	0.00	0.00	0.09	0.81	0.09	0.00	0.00
	2	0.00	0.01	0.08	0.65	0.25	0.01	0.00
	3	0.00	0.01	0.08	0.82	0.09	0.00	0.00
	4	0.00	0.00	0.03	0.74	0.21	0.01	0.00
	5	0.00	0.00	0.07	0.79	0.13	0.01	0.00
33	Overall	0.00	0.00	0.04	0.90	0.05	0.00	0.00
	1	0.00	0.00	0.04	0.90	0.05	0.00	0.00
	2	0.00	0.01	0.05	0.91	0.03	0.00	0.00
	3	0.00	0.00	0.04	0.92	0.04	0.00	0.00
	4	0.00	0.00	0.02	0.90	0.07	0.00	0.00
	5	0.00	0.00	0.04	0.90	0.05	0.00	0.00
34	Overall	0.00	0.00	0.06	0.76	0.17	0.00	0.00
	1	0.00	0.00	0.07	0.78	0.14	0.00	0.00
	2	0.00	0.01	0.07	0.76	0.15	0.01	0.00
	3	0.00	0.00	0.04	0.79	0.16	0.01	0.00
	4	0.00	0.00	0.05	0.77	0.18	0.00	0.00
	5	0.00	0.00	0.05	0.73	0.21	0.00	0.00
35	Overall	0.00	0.00	0.03	0.94	0.02	0.00	0.00
	1	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	2	0.00	0.00	0.01	0.96	0.02	0.00	0.00
	3	0.00	0.00	0.03	0.95	0.02	0.00	0.00
	4	0.00	0.00	0.04	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.94	0.02	0.00	0.00

**Table F-5. NYS Public Schools Grade 7 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
31	Overall	0.00	0.00	0.04	0.92	0.03	0.00	0.00
	1	0.00	0.00	0.04	0.91	0.04	0.00	0.00
	2	0.00	0.01	0.08	0.87	0.03	0.01	0.00
	3	0.00	0.00	0.04	0.92	0.03	0.00	0.00
	4	0.00	0.01	0.04	0.91	0.04	0.00	0.00
	5	0.00	0.00	0.04	0.93	0.03	0.00	0.00
32	Overall	0.00	0.00	0.05	0.86	0.08	0.00	0.00
	1	0.00	0.00	0.04	0.86	0.10	0.00	0.00
	2	0.00	0.00	0.01	0.90	0.09	0.00	0.00
	3	0.00	0.00	0.06	0.87	0.06	0.00	0.00
	4	0.00	0.00	0.07	0.85	0.08	0.00	0.00
	5	0.00	0.00	0.06	0.87	0.07	0.00	0.00
33	Overall	0.00	0.00	0.05	0.88	0.07	0.00	0.00
	1	0.00	0.00	0.05	0.88	0.06	0.00	0.00
	2	0.00	0.01	0.06	0.91	0.01	0.01	0.00
	3	0.00	0.00	0.04	0.90	0.06	0.00	0.00
	4	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	5	0.00	0.00	0.04	0.88	0.08	0.00	0.00
34	Overall	0.00	0.00	0.13	0.74	0.13	0.00	0.00
	1	0.00	0.00	0.11	0.76	0.13	0.00	0.00
	2	0.00	0.00	0.08	0.75	0.15	0.02	0.00
	3	0.00	0.00	0.16	0.72	0.11	0.00	0.00
	4	0.00	0.01	0.17	0.69	0.13	0.00	0.00
	5	0.00	0.00	0.13	0.72	0.13	0.00	0.00
35	Overall	0.00	0.00	0.06	0.85	0.08	0.00	0.00
	1	0.00	0.00	0.05	0.86	0.08	0.00	0.00
	2	0.00	0.01	0.11	0.76	0.12	0.00	0.00
	3	0.00	0.00	0.06	0.88	0.06	0.00	0.00
	4	0.00	0.00	0.09	0.84	0.06	0.00	0.00
	5	0.00	0.00	0.07	0.83	0.09	0.01	0.00
36	Overall	0.00	0.01	0.12	0.77	0.10	0.01	0.00
	1	0.00	0.01	0.13	0.77	0.08	0.00	0.00
	2	0.00	0.03	0.24	0.65	0.05	0.04	0.01
	3	0.00	0.01	0.07	0.75	0.16	0.01	0.00
	4	0.00	0.01	0.14	0.77	0.08	0.00	0.00
	5	0.00	0.01	0.10	0.77	0.11	0.01	0.00

**Table F-5. NYS Public Schools Grade 7 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
37	Overall	0.00	0.00	0.05	0.86	0.08	0.01	0.00
	1	0.00	0.00	0.04	0.87	0.08	0.01	0.00
	2	0.01	0.00	0.07	0.79	0.12	0.01	0.00
	3	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	4	0.00	0.01	0.08	0.86	0.05	0.01	0.00
	5	0.00	0.00	0.05	0.85	0.08	0.01	0.00
38	Overall	0.00	0.00	0.04	0.73	0.22	0.00	0.00
	1	0.00	0.00	0.04	0.76	0.19	0.00	0.00
	2	0.00	0.00	0.06	0.70	0.24	0.00	0.00
	3	0.00	0.00	0.04	0.70	0.25	0.01	0.00
	4	0.00	0.00	0.03	0.73	0.23	0.00	0.00
	5	0.00	0.00	0.03	0.71	0.25	0.00	0.00

**Table F-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
28	Overall	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	1	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	2	0.00	0.00	0.01	0.99	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.97	0.02	0.00	0.00
	4	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	5	0.00	0.00	0.01	0.97	0.01	0.00	0.00
29	Overall	0.00	0.00	0.06	0.67	0.27	0.01	0.00
	1	0.00	0.00	0.05	0.65	0.29	0.01	0.00
	2	0.00	0.00	0.03	0.68	0.29	0.00	0.00
	3	0.00	0.00	0.08	0.70	0.21	0.01	0.00
	4	0.00	0.00	0.04	0.70	0.26	0.00	0.00
	5	0.00	0.00	0.05	0.68	0.26	0.00	0.00
30	Overall	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	2	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	3	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	4	0.00	0.00	0.01	0.96	0.02	0.00	0.00
	5	0.00	0.00	0.01	0.95	0.03	0.00	0.00
31	Overall	0.00	0.00	0.05	0.89	0.06	0.00	0.00
	1	0.00	0.00	0.05	0.88	0.06	0.00	0.00
	2	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	3	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	4	0.00	0.00	0.04	0.90	0.05	0.00	0.00
	5	0.00	0.00	0.05	0.90	0.05	0.00	0.00
32	Overall	0.00	0.01	0.05	0.86	0.07	0.01	0.00
	1	0.00	0.01	0.06	0.86	0.07	0.01	0.00
	2	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	3	0.00	0.01	0.04	0.87	0.07	0.01	0.00
	4	0.00	0.01	0.04	0.87	0.07	0.01	0.00
	5	0.00	0.00	0.04	0.87	0.08	0.01	0.00
33	Overall	0.00	0.00	0.08	0.73	0.17	0.01	0.00
	1	0.00	0.01	0.09	0.72	0.16	0.02	0.00
	2	0.00	0.01	0.06	0.74	0.19	0.01	0.00
	3	0.00	0.00	0.08	0.77	0.15	0.01	0.00
	4	0.00	0.00	0.04	0.74	0.20	0.02	0.00
	5	0.00	0.00	0.08	0.71	0.19	0.01	0.00

**Table F-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
34	Overall	0.00	0.00	0.10	0.82	0.08	0.00	0.00
	1	0.00	0.00	0.11	0.80	0.08	0.00	0.00
	2	0.00	0.00	0.07	0.82	0.11	0.00	0.00
	3	0.00	0.00	0.09	0.83	0.08	0.00	0.00
	4	0.00	0.01	0.04	0.87	0.08	0.00	0.00
	5	0.00	0.00	0.10	0.84	0.07	0.00	0.00
35	Overall	0.00	0.00	0.08	0.86	0.05	0.00	0.00
	1	0.00	0.01	0.10	0.85	0.05	0.00	0.00
	2	0.00	0.00	0.08	0.87	0.04	0.01	0.00
	3	0.00	0.00	0.08	0.86	0.06	0.01	0.00
	4	0.00	0.00	0.08	0.87	0.04	0.00	0.00
	5	0.00	0.00	0.06	0.88	0.06	0.00	0.00
36	Overall	0.00	0.00	0.06	0.89	0.05	0.00	0.00
	1	0.00	0.00	0.07	0.88	0.05	0.00	0.00
	2	0.00	0.00	0.07	0.89	0.04	0.00	0.00
	3	0.00	0.00	0.05	0.89	0.05	0.00	0.00
	4	0.00	0.00	0.06	0.88	0.05	0.00	0.00
	5	0.00	0.00	0.04	0.90	0.05	0.00	0.00
37	Overall	0.00	0.01	0.13	0.79	0.07	0.00	0.00
	1	0.00	0.01	0.14	0.78	0.07	0.00	0.00
	2	0.00	0.01	0.06	0.82	0.10	0.01	0.00
	3	0.00	0.01	0.13	0.77	0.09	0.00	0.00
	4	0.00	0.01	0.11	0.80	0.08	0.00	0.00
	5	0.00	0.01	0.12	0.81	0.07	0.00	0.00
38	Overall	0.00	0.00	0.11	0.83	0.06	0.00	0.00
	1	0.00	0.00	0.11	0.82	0.06	0.00	0.00
	2	0.00	0.01	0.10	0.87	0.03	0.00	0.00
	3	0.00	0.01	0.12	0.81	0.07	0.00	0.00
	4	0.00	0.01	0.13	0.80	0.05	0.00	0.00
	5	0.00	0.00	0.07	0.86	0.06	0.00	0.00
39	Overall	0.00	0.00	0.10	0.76	0.13	0.00	0.00
	1	0.00	0.00	0.09	0.76	0.14	0.00	0.00
	2	0.00	0.00	0.13	0.81	0.06	0.00	0.00
	3	0.00	0.01	0.11	0.76	0.13	0.00	0.00
	4	0.00	0.01	0.07	0.82	0.10	0.00	0.00
	5	0.00	0.00	0.10	0.74	0.15	0.00	0.00

**Table F-6. NYS Public Schools Grade 8 Mathematics Operational Test 2009:
Percentages of Score Differences [Local Score Minus Audit Score]
(continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
40	Overall	0.00	0.01	0.13	0.77	0.09	0.00	0.00
	1	0.00	0.01	0.15	0.75	0.09	0.00	0.00
	2	0.00	0.01	0.07	0.88	0.03	0.01	0.00
	3	0.00	0.01	0.14	0.75	0.09	0.00	0.00
	4	0.00	0.01	0.08	0.78	0.13	0.01	0.00
	5	0.00	0.01	0.09	0.82	0.09	0.00	0.00
41	Overall	0.00	0.00	0.04	0.88	0.07	0.00	0.00
	1	0.00	0.00	0.05	0.92	0.04	0.00	0.00
	2	0.00	0.00	0.04	0.94	0.02	0.00	0.00
	3	0.00	0.00	0.03	0.84	0.12	0.00	0.00
	4	0.00	0.00	0.03	0.90	0.07	0.00	0.00
	5	0.00	0.00	0.04	0.83	0.13	0.00	0.00
42	Overall	0.00	0.01	0.07	0.83	0.09	0.00	0.00
	1	0.00	0.01	0.07	0.83	0.09	0.00	0.00
	2	0.00	0.02	0.08	0.84	0.06	0.00	0.00
	3	0.00	0.00	0.09	0.83	0.07	0.00	0.00
	4	0.00	0.01	0.07	0.81	0.11	0.00	0.00
	5	0.00	0.01	0.07	0.82	0.10	0.00	0.00
43	Overall	0.00	0.00	0.11	0.75	0.13	0.01	0.00
	1	0.00	0.01	0.12	0.75	0.12	0.01	0.00
	2	0.00	0.01	0.11	0.76	0.11	0.02	0.00
	3	0.00	0.00	0.10	0.76	0.13	0.01	0.00
	4	0.00	0.01	0.13	0.74	0.12	0.01	0.00
	5	0.00	0.00	0.08	0.73	0.18	0.01	0.00
44	Overall	0.00	0.01	0.08	0.83	0.07	0.01	0.00
	1	0.00	0.01	0.08	0.83	0.07	0.01	0.00
	2	0.00	0.03	0.07	0.84	0.05	0.01	0.00
	3	0.00	0.02	0.08	0.82	0.07	0.01	0.00
	4	0.00	0.01	0.07	0.84	0.05	0.01	0.00
	5	0.00	0.01	0.08	0.82	0.08	0.02	0.00
45	Overall	0.00	0.01	0.07	0.84	0.08	0.01	0.00
	1	0.00	0.00	0.08	0.83	0.07	0.01	0.00
	2	0.00	0.00	0.05	0.91	0.04	0.01	0.00
	3	0.00	0.01	0.08	0.83	0.07	0.01	0.00
	4	0.00	0.01	0.05	0.85	0.08	0.00	0.00
	5	0.00	0.01	0.06	0.84	0.09	0.01	0.00

Appendix G

Item Level Differences for Mathematics Including All Schools in State

Without New York City Schools

**Table G-1. NYS Public Schools (Without NYC) Grade 3 Mathematics
Operational Test 2009: Percentages of Score Differences
[Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
26	Overall	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.01	0.96	0.02	0.00	0.00
	2	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	3	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	4	0.00	0.00	0.01	0.95	0.03	0.00	0.00
	5	0.00	0.00	0.02	0.94	0.03	0.00	0.00
27	Overall	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	1	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	2	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	3	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	4	0.00	0.00	0.01	0.97	0.01	0.00	0.00
	5	0.00	0.00	0.02	0.96	0.01	0.00	0.00
28	Overall	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	1	0.00	0.00	0.06	0.90	0.03	0.00	0.00
	2	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	3	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	4	0.00	0.00	0.07	0.89	0.04	0.00	0.00
	5	0.00	0.00	0.05	0.92	0.04	0.00	0.00
29	Overall	0.00	0.00	0.05	0.89	0.06	0.00	0.00
	1	0.00	0.00	0.04	0.90	0.05	0.00	0.00
	2	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	3	0.00	0.00	0.05	0.89	0.06	0.00	0.00
	4	0.00	0.00	0.06	0.89	0.05	0.00	0.00
	5	0.00	0.00	0.05	0.88	0.06	0.00	0.00
30	Overall	0.00	0.01	0.06	0.91	0.03	0.00	0.00
	1	0.00	0.01	0.07	0.90	0.03	0.00	0.00
	2	0.00	0.01	0.04	0.92	0.03	0.00	0.00
	3	0.00	0.01	0.06	0.91	0.03	0.00	0.00
	4	0.00	0.00	0.06	0.91	0.02	0.00	0.00
	5	0.00	0.01	0.06	0.90	0.03	0.00	0.00
31	Overall	0.00	0.00	0.08	0.88	0.04	0.00	0.00
	1	0.00	0.00	0.09	0.86	0.05	0.00	0.00
	2	0.00	0.00	0.03	0.94	0.02	0.00	0.00
	3	0.00	0.00	0.08	0.88	0.04	0.00	0.00
	4	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	5	0.00	0.00	0.09	0.88	0.03	0.00	0.00

**Table G-2. NYS Public Schools (Without NYC) Grade 4 Mathematics
Operational Test 2009: Percentages of Score Differences
[Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
31	Overall	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	1	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	2	0.00	0.00	0.00	0.83	0.15	0.02	0.00
	3	0.00	0.00	0.05	0.90	0.05	0.00	0.00
	4	0.00	0.00	0.06	0.89	0.05	0.01	0.00
	5	0.00	0.00	0.03	0.89	0.07	0.00	0.00
32	Overall	0.00	0.00	0.01	0.99	0.01	0.00	0.00
	1	0.00	0.00	0.00	0.99	0.01	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.99	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.98	0.01	0.00	0.00
	5	0.00	0.00	0.01	0.99	0.01	0.00	0.00
33	Overall	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	1	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.03	0.95	0.02	0.00	0.00
	4	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.04	0.00	0.00
34	Overall	0.00	0.00	0.04	0.93	0.03	0.00	0.00
	1	0.00	0.00	0.06	0.91	0.03	0.00	0.00
	2	0.00	0.00	0.07	0.93	0.00	0.00	0.00
	3	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.93	0.04	0.00	0.00
	5	0.00	0.00	0.05	0.93	0.02	0.00	0.00
35	Overall	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	1	0.00	0.00	0.06	0.89	0.05	0.00	0.00
	2	0.00	0.00	0.04	0.96	0.00	0.00	0.00
	3	0.00	0.00	0.06	0.90	0.04	0.00	0.00
	4	0.00	0.00	0.05	0.89	0.06	0.00	0.00
	5	0.00	0.00	0.06	0.89	0.04	0.00	0.00
36	Overall	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	1	0.00	0.00	0.03	0.91	0.06	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.92	0.05	0.00	0.00
	5	0.00	0.00	0.04	0.90	0.06	0.00	0.00

**Table G-2. NYS Public Schools (Without NYC) Grade 4 Mathematics
Operational Test 2009: Percentages of Score Differences
[Local Score Minus Audit Score] (continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
37	Overall	0.00	0.00	0.04	0.92	0.04	0.00	0.00
	1	0.00	0.00	0.04	0.90	0.05	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.03	0.93	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	5	0.00	0.00	0.04	0.92	0.04	0.00	0.00
38	Overall	0.00	0.00	0.10	0.73	0.16	0.00	0.00
	1	0.00	0.00	0.13	0.71	0.15	0.00	0.00
	2	0.00	0.00	0.04	0.70	0.26	0.00	0.00
	3	0.00	0.00	0.09	0.74	0.16	0.00	0.00
	4	0.00	0.00	0.10	0.74	0.15	0.00	0.00
	5	0.00	0.00	0.09	0.72	0.18	0.00	0.00
39	Overall	0.00	0.00	0.02	0.94	0.05	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.00	0.02	0.91	0.07	0.00	0.00
	3	0.00	0.00	0.01	0.94	0.04	0.00	0.00
	4	0.00	0.00	0.02	0.96	0.03	0.00	0.00
	5	0.00	0.00	0.02	0.91	0.07	0.00	0.00
40	Overall	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	1	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	4	0.00	0.00	0.01	0.98	0.01	0.00	0.00
	5	0.00	0.00	0.01	0.98	0.01	0.00	0.00
41	Overall	0.00	0.00	0.04	0.90	0.06	0.00	0.00
	1	0.00	0.00	0.03	0.88	0.09	0.00	0.00
	2	0.00	0.00	0.04	0.96	0.00	0.00	0.00
	3	0.00	0.00	0.04	0.90	0.06	0.00	0.00
	4	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	5	0.00	0.00	0.04	0.91	0.05	0.00	0.00
42	Overall	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	1	0.00	0.00	0.03	0.95	0.02	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.97	0.02	0.00	0.00
	4	0.00	0.00	0.02	0.97	0.01	0.00	0.00
	5	0.00	0.00	0.02	0.96	0.02	0.00	0.00

**Table G-2. NYS Public Schools (Without NYC) Grade 4 Mathematics
Operational Test 2009: Percentages of Score Differences
[Local Score Minus Audit Score] (continued)**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
43	Overall	0.00	0.00	0.01	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.02	0.02	0.96	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.96	0.03	0.00	0.00
	4	0.00	0.00	0.00	0.95	0.05	0.00	0.00
	5	0.00	0.00	0.01	0.95	0.03	0.00	0.00
44	Overall	0.00	0.00	0.02	0.94	0.03	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.04	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.02	0.94	0.03	0.00	0.00
	4	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.04	0.00	0.00
45	Overall	0.00	0.00	0.02	0.95	0.02	0.00	0.00
	1	0.00	0.00	0.02	0.95	0.02	0.00	0.00
	2	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	3	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	4	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	5	0.00	0.00	0.02	0.94	0.03	0.00	0.00
46	Overall	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	1	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	2	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	3	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	4	0.00	0.00	0.01	0.93	0.06	0.00	0.00
	5	0.00	0.00	0.04	0.90	0.06	0.00	0.00
47	Overall	0.00	0.00	0.03	0.90	0.06	0.00	0.00
	1	0.00	0.00	0.03	0.91	0.06	0.00	0.00
	2	0.00	0.00	0.02	0.98	0.00	0.00	0.00
	3	0.00	0.00	0.04	0.91	0.05	0.00	0.00
	4	0.00	0.00	0.02	0.87	0.10	0.00	0.00
	5	0.00	0.00	0.04	0.89	0.07	0.00	0.00
48	Overall	0.00	0.00	0.02	0.95	0.03	0.00	0.00
	1	0.00	0.00	0.03	0.94	0.03	0.00	0.00
	2	0.00	0.00	0.00	1.00	0.00	0.00	0.00
	3	0.00	0.00	0.02	0.96	0.02	0.00	0.00
	4	0.00	0.00	0.02	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.94	0.03	0.00	0.00

**Table G-3. NYS Public Schools (Without NYC) Grade 5 Mathematics
Operational Test 2009: Percentages of Score Differences
[Local Score Minus Audit Score]**

Item #	Scoring Model	-3	-2	-1	0	1	2	3
27	Overall	0.00	0.00	0.06	0.90	0.03	0.00	0.00
	1	0.00	0.00	0.08	0.90	0.02	0.00	0.00
	2	0.00	0.00	0.06	0.92	0.01	0.00	0.00
	3	0.00	0.00	0.06	0.91	0.03	0.00	0.00
	4	0.00	0.00	0.04	0.93	0.03	0.00	0.00
	5	0.00	0.00	0.06	0.90	0.04	0.00	0.00
28	Overall	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	1	0.00	0.00	0.04	0.92	0.04	0.00	0.00
	2	0.00	0.00	0.04	0.92	0.04	0.00	0.00
	3	0.00	0.00	0.03	0.93	0.04	0.00	0.00
	4	0.00	0.00	0.03	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.03	0.00	0.00
29	Overall	0.00	0.00	0.02	0.96	0.03	0.00	0.00
	1	0.00	0.00	0.02	0.94	0.03	0.00	0.00
	2	0.00	0.00	0.01	0.95	0.04	0.00	0.00
	3	0.00	0.00	0.01	0.96	0.02	0.00	0.00
	4	0.00	0.00	0.01	0.97	0.02	0.00	0.00
	5	0.00	0.00	0.02	0.95	0.03	0.00	0.00
30	Overall	0.00	0.00	0.03	0.92	0.04	0.00	0.00
	1	0.00	0.00	0.03	0.92	0.05	0.00	0.00
	2	0.00	0.00	0.02	0.92	0.05	0.01	0.00
	3	0.00	0.00	0.03	0.92	0.04	0.00	0.00
	4	0.00	0.00	0.02	0.95	0.02	0.00	0.00
	5	0.00	0.00	0.03	0.93	0.04	0.00	0.00
31	Overall	0.00	0.01	0.05	0.82	0.12	0.01	0.00
	1	0.00	0.01	0.05	0.79	0.13	0.01	0.00
	2	0.00	0.00	0.05	0.86	0.08	0.00	0.00
	3	0.00	0.00	0.04	0.80	0.14	0.01	0.00
	4	0.00	0.00	0.03	0.88	0.08	0.00	0.00
	5	0.00	0.01	0.05	0.83	0.11	0.00	0.00
32	Overall	0.00	0.01	0.12	0.68	0.18	0.01	0.00
	1	0.00	0.00	0.10	0.74	0.14	0.01	0.00
	2	0.00	0.00	0.07	0.70	0.23	0.00	0.00
	3	0.00	0.01	0.14	0.67	0.17	0.01	0.00
	4	0.00	0.00	0.18	0.72	0.09	0.00	0.00
	5	0.00	0.01	0.10	0.65	0.23	0.01	0.00