

Inter-Rater Reliability Study of New York State
Grades 3-8 English Language Arts Test
2007 Test Administration

Technical Report

Prepared by:
Pearson
February 7, 2008

Table of Contents

I. Purpose and Scope of Audit	1
II. Selection of School Sample and Test Books	3
III. Data Collection and School Participation	6
IV. Selection and Training of Auditors	8
V. Audit Procedures	10
VI. Data Analysis	11
VII. Results	13
VIII. Summary	20
Appendix A: ELA Item Analysis for Local Scoring	21
Appendix B: ELA Item Analysis for Audit Scoring	23
Appendix C: Item Level Statistics for ELA Including All Schools in State	25
Appendix D: Item Level Statistics for ELA Including All Schools in State Without New York City Schools.....	32
Appendix E: Item Level Statistics for ELA Including New York City Schools Only	39
Appendix F: Item Level Differences for ELA Including All Schools in State.....	41
Appendix G: Item Level Differences for ELA All Schools in State Without New York City Schools.....	48
Appendix H: Item Level Differences for ELA Including New York City Schools Only.....	55

Table of Contents Continued

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 ELA	6
Table 8:	Obtained Proportions for ELA	7
Table 9:	New York State Public Schools ELA Operational Test 2007: Inter-rater Agreement.....	14
Table 10:	Percentage of Raw Score Differences for ELA (Audit Scoring Minus Local Scoring).....	17
Table 11:	Correlations Between Local and Audit Scores.....	18
Table C-1:	New York State Public Schools Grade 3 ELA Operational Test 2007: Inter-rater Agreement	26
Table C-2:	New York State Public Schools Grade 4 ELA Operational Test 2007: Inter-rater Agreement	27
Table C-3:	New York State Public Schools Grade 5 ELA Operational Test 2007: Inter-rater Agreement	28
Table C-4:	New York State Public Schools Grade 6 ELA Operational Test 2007: Inter-rater Agreement	29
Table C-5:	New York State Public Schools Grade 7 ELA Operational Test 2007: Inter-rater Agreement	30

List of Tables Continued

Table C-6:	New York State Public Schools Grade 8 ELA Operational Test 2007: Inter-rater Agreement	31
Table D-1:	New York State Public Schools (Without NYC) Grade 3 ELA Operational Test 2007: Inter-rater Agreement	33
Table D-2:	New York State Public Schools (Without NYC) Grade 4 ELA Operational Test 2007: Inter-rater Agreement	34
Table D-3:	New York State Public Schools (Without NYC) Grade 5 ELA Operational Test 2007: Inter-rater Agreement	35
Table D-4:	New York State Public Schools (Without NYC) Grade 6 ELA Operational Test 2007: Inter-rater Agreement	36
Table D-5:	New York State Public Schools (Without NYC) Grade 7 ELA Operational Test 2007: Inter-rater Agreement	37
Table D-6:	New York State Public Schools (Without NYC) Grade 8 ELA Operational Test 2007: Inter-rater Agreement	38
Table E-1:	NYC Public Schools 3-8 ELA Operational Test 2007: Inter-rater Agreement	40
Table F-1:	New York State Public Schools Grade 3 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	42
Table F-2:	New York State Public Schools Grade 4 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	43
Table F-3:	New York State Public Schools Grade 5 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	44
Table F-4:	New York State Public Schools Grade 6 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	45
Table F-5:	New York State Public Schools Grade 7 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	46

List of Tables Continued

Table F-6:	New York State Public Schools Grade 8 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	47
Table G-1:	New York State Public Schools Grade 3 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	49
Table G-2:	New York State Public Schools Grade 4 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	50
Table G-3:	New York State Public Schools Grade 5 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	51
Table G-4:	New York State Public Schools Grade 6 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	52
Table G-5:	New York State Public Schools Grade 7 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	53
Table G-6:	New York State Public Schools Grade 8 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	54
Table H-1:	New York State Public Schools Grade 3-8 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring].....	56

I. Purpose and Scope of Audit

Purpose

The New York State Grades 3-8 English Language Arts (ELA) assessments consist of both multiple-choice (MC) and constructed-response (CR) items. The multiple-choice items are scored at the Regional Information Centers across the state and the constructed-response items are scored by teachers at the regional scoring centers or in their districts or 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, the Department conducts annual scoring audits that involve independent rescoring of five percent of all test books 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 books 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 books for each score point using scoring rubrics for the constructed-response questions; teachers have consistently done a very good job scoring the state assessment books.

Schools identified for the 2007 audit were instructed to send their student assessments to Pearson for rescoring. Pearson is a professional scoring company known throughout the country for its 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 ELA assessments were administered in January 2007 throughout the state. For the first time, the operational data for these assessments were collected by NYSED, including both MC and CR scores. The Regional Information Centers scored the MC items and New York teachers scored the CR items. In March 2007, Pearson conducted the audit study by rescoring the CR items from approximately five percent of all test books. Pearson identified a stratified sample of schools (about 180 schools per grade) from across the state for each of the grade levels that contained approximately 15,000 student test books. 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 75,185 ELA test books were collected from sample schools and rescored in March 2007.

Audit notification letters were sent to the sample schools in February 2007 and the selected schools sent their student test books to Pearson for audit. Pearson rescored the constructed-response 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 teachers, and the second set came from the audit scoring performed by Pearson. The data analysis performed in this study consisted of various comparisons between the local scores and the audit scores.

II. Selection of School Sample and Test Books

Audit Samples

To achieve the target audit sample of 12,500 test books per grade level, approximately 15,000 test books were sampled. Six stratified random samples of schools, with approximately 180 schools per grade, were selected, one for each grade, from all New York schools with grade 3-8 enrollment to yield the target number of test books. Each school was selected for audit at only one grade level. All selected schools were requested to send Pearson their ELA test books 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 N/RC 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 N/RC Districts		All districts between the 20 th and 70 th percentiles on the index
Low N/RC 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/RC 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
Total	195664	196079	202035	204587	212128	213514	1224007
New York City	68648	67190	69443	69356	70905	71714	417256
Large Cities	8173	7939	8234	8542	10008	9219	52115
High Need Urban/Suburban	15395	15432	15438	16171	16622	16499	95557
High Need Rural	11490	11457	11780	12665	13545	13580	74517
Average Need	59684	61185	63416	65379	68944	70456	389064
Low Need	30692	31686	32103	31551	31316	31446	188794
Charter	1582	1190	1621	923	788	600	6704

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							
N/RC Category	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
New York City	0.35	0.34	0.34	0.34	0.33	0.34	0.34
Large Cities	0.04	0.04	0.04	0.04	0.05	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	0.33	0.33	0.32
Low Need	0.16	0.16	0.16	0.15	0.15	0.15	0.15
Charter	0.01	0.01	0.01	0.00	0.00	0.00	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	5263	5140	5156	5085	5014	5038	30696
Large Cities	627	607	611	626	708	648	3827
High Need Urban/Suburban	1180	1181	1146	1186	1175	1159	7027
High Need Rural	881	876	875	929	958	954	5472
Average Need	4575	4681	4708	4793	4875	4950	28583
Low Need	2353	2424	2383	2313	2214	2209	13897
Charter	121	91	120	68	56	42	498
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	5287	5116	5175	5088	5070	5071	30807
Large Cities	639	611	613	607	723	644	3837
High Need Urban/Suburban	1180	1194	1162	1208	1180	1135	7059
High Need Rural	881	879	879	922	967	962	5490
Average Need	4585	4709	4708	4824	4855	4969	28650
Low Need	2341	2428	2411	2339	2250	2216	13985
Charter	128	94	127	67	74	50	540
Totals	15041	15031	15075	15055	15119	15047	90368

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'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	0.35	0.34	0.34	0.34	0.34	0.34	0.34
Large Cities	0.04	0.04	0.04	0.04	0.05	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	0.30	0.31	0.31	0.32	0.32	0.33	0.32
Low Need	0.16	0.16	0.16	0.16	0.15	0.15	0.15
Charter	0.01	0.01	0.01	0.00	0.00	0.00	0.01

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

III. Data Collection and School Participation

Pearson notified 764 schools, 705 of which returned materials. This represents a participation rate of 92%.

After the test books 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 ELA

Target Proportions						
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
New York City	5176	4954	4874	4432	4172	3671
Large Cities	598	618	413	469	617	657
High Need Urban/Suburban	1059	1065	743	1066	638	1077
High Need Rural	720	859	812	554	905	990
Average Need	3891	4062	3851	4220	3760	4082
Low Need	1755	1532	2106	1498	1118	1862
Charter	88	67	0	60	48	46

Table 8. Obtained Proportions for ELA

Target Proportions						
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
New York City	0.39	0.38	0.38	0.36	0.37	0.30
Large Cities	0.05	0.05	0.03	0.04	0.05	0.05
High Need Urban/Suburban	0.08	0.08	0.06	0.09	0.06	0.09
High Need Rural	0.05	0.07	0.06	0.05	0.08	0.08
Average Need	0.29	0.31	0.30	0.34	0.33	0.33
Low Need	0.13	0.12	0.16	0.12	0.10	0.15
Charter	0.01	0.01	0.00	0.00	0.00	0.00

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 for the grade 7 Low Need group, given the actual sample was 5% smaller than anticipated. All other samples differed less than 5% from the targets.

IV. Selection and Training of Auditors

A. 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 Web site 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. Further, ELA supervisors on grades 4, 6, and 8 were required to demonstrate strong grammar skills.

Auditors possessed, at a minimum, a four-year college degree. They were placed on the most appropriate subject area based on their educational qualifications and their work or scoring experience. Auditors who demonstrated strong grammar skills in a grammar placement test qualified to assign mechanics scores to linked items.

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

B. Training of auditors

Steps used to support that trainees were adequately trained

Supervisor training in Columbus and Atlanta took place from March 14–16, 2007. Supervisors trained on all books and all grades for which they would score. Auditors (48 in Columbus and 168 in Atlanta) began training on March 19, 2007. Auditors trained on items in a single book, completed scoring all books, and then trained on a new book for the next grade level.

Pearson staff used only those training materials supplied by the NYSED and used in the original scorer training. Scoring directors began training by reviewing and discussing the scoring guides for items in a book. 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 books. Subsequent practice sets for a book 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 books and any necessary review, auditors completed the consistency assurance set (CAS) for that book. A review and discussion of the scores occurred after auditors had assigned scores to all books 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 books. Trainees who were unable to demonstrate accuracy and consistency in scoring were not allowed to score the project.

C. 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 were:

- **Backreading (read behinds)** was one of the primary responsibilities of scoring directors and scoring supervisors and began immediately. 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** begins immediately with 10% of responses in the audit receiving an independent score by a different auditor than the original. Second score books are randomly generated by the system. By having a different auditor score the book a second time without knowledge of the score given by the original auditor, it generates 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 Atlanta (ELA grades 4, 6, and 8), auditors were divided into two groups per grade. One group scored Book 2 only. The second group of auditors scored all of Book 3 and assigned a mechanics score to the linked items in Books 2 and 3.

In Columbus, auditors for ELA were also divided into two groups per grade. Each group scored either Book 1 or Book 2 for grades 3, 5, and 7.

At both sites completed scoring monitors were 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, 21 ELA 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 scoring. An intra-class correlation was computed between the local and audit scoring which provides an estimate of the reliability of the scoring. 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 scoring 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 question 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 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. 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 constructed-response question, 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 results was computed. The frequency of these differences was 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 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.

VII. Results

Item Means

The average score for each constructed-response question was computed based on the local scoring and the audit scoring as well as the standard deviation. The results from this analysis are presented in Table 9. They show a very close agreement between the local scoring and the audit scoring on the ELA constructed-response questions. 7 out of 21 items have exactly the same mean raw scores and an additional 8 items have a mean difference of 0.1.

Table 9. New York State Public Schools ELA Operational Test 2007: Inter-rater Agreement

Grade	Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
					Exact	Approx.	Total	Local	Audit	Local	Audit		
3	21	Overall	2	12921	95.2	4.5	99.7	1.3	1.3	0.86	0.87	0.98	0.93
	26	Overall	2	12921	86.1	13.2	99.3	1.8	1.8	0.46	0.46	0.76	0.86
	27	Overall	2	12921	88.6	10.6	99.2	1.6	1.6	0.69	0.71	0.92	0.89
	28	Overall	3	12921	93.6	5.3	98.9	2.8	2.8	0.69	0.70	0.93	0.94
4	29	Overall	4	13014	48.7	46.5	95.2	2.8	2.5	0.87	0.87	0.73	0.42
	30	Overall	4	13014	53.0	43.9	96.9	2.7	2.6	0.85	0.85	0.76	0.53
	31	Overall	3	13014	53.7	42.9	96.6	2.1	2.0	0.74	0.80	0.69	0.54
5	21	Overall	2	12536	77.3	21.7	99.0	1.4	1.4	0.75	0.78	0.88	0.61
	26	Overall	2	12536	76.9	22.2	99.1	1.3	1.2	0.67	0.68	0.84	0.77
	27	Overall	3	12536	73.4	25.0	98.4	1.3	1.4	1.05	1.04	0.92	0.73
6	27	Overall	5	11855	38.9	48.5	87.4	3.5	3.0	1.00	0.95	0.65	0.37
	28	Overall	5	11855	39.9	47.2	87.1	3.3	2.9	1.07	1.08	0.72	0.40
	29	Overall	3	11855	54.5	42.8	97.4	2.2	2.0	0.70	0.73	0.66	0.55
7	3	Overall	2	11055	87.6	11.8	99.5	1.7	1.6	0.60	0.65	0.90	0.71
	22	Overall	2	11055	64.7	34.7	99.3	1.5	1.2	0.64	0.66	0.73	0.65
	33	Overall	2	11055	70.0	27.9	97.9	1.5	1.5	0.61	0.67	0.71	0.70
	34	Overall	2	11055	75.4	23.9	99.3	1.7	1.6	0.57	0.57	0.74	0.75
	35	Overall	3	11055	69.4	28.1	97.4	1.0	1.0	0.92	0.90	0.87	0.69
8	27	Overall	5	12245	45.3	46.3	91.6	3.8	3.7	1.10	1.03	0.78	0.45
	28	Overall	5	12245	42.5	46.8	89.3	3.6	3.3	1.16	1.12	0.78	0.42
	29	Overall	3	12245	57.4	40.2	97.6	2.3	2.2	0.70	0.74	0.69	0.57

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.

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 scoring 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 5 maximum points. Taken collectively, the total percent of agreement ranges from a low of 87.1% to a high of 99.7%. Consistent with the information in the item means, the percent of agreement shows a high level of agreement between the local and audit scoring.

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 reliability of the scoring based on an average of the local and audit scores.

Generally, correlations greater than 0.60 are considered very strong because they explain more than one-third of the variance. Table 9 shows that all of the items had correlations greater than 0.60. One-third of the items had correlations greater than 0.80. The intra-class correlations range from 0.65 to 0.98. The intra-class correlations show a high degree of consistency between the local and audit scores.

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:

- $k > .75$ denotes excellent reproducibility
- $0.4 \leq k \leq .75$ 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 scoring. Item 27 in grade 6 produced a weighted kappa statistic of 0.37 which is slightly less than the category of “good reproducibility”. All other items had weighted kappa statistics denoted good or excellent reproducibility.

Inter-rater Agreement

For each constructed-response question, 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 results were 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 Raw Score Differences for English (Audit Scoring Minus Local Scoring)

Grade	Item	MAX Points	Difference				
			0	1	2	3	4
3	21	2	95%	5%	0%		
N=12921	26	2	86%	13%	0%		
	27	2	89%	10%	1%		
	28	3	94%	5%	0%	0%	
4	29	4	49%	47%	5%	0%	0%
N=13014	30	4	53%	44%	3%	0%	
	31	3	54%	43%	3%	0%	
5	21	2	77%	22%	1%		
N=12536	26	2	77%	22%	1%		
	27	3	73%	25%	2%	0%	
6	27	5	39%	48%	12%	1%	0%
N=11855	28	5	40%	47%	12%	1%	0%
	29	3	55%	43%	2%	0%	
7	3	2	88%	12%	0%		
N=11055	22	2	65%	35%	1%		
	33	2	70%	28%	2%		
	34	2	75%	24%	0%		
	35	3	69%	28%	2%	0%	
8	27	5	45%	46%	8%	0%	0%
N=12245	28	5	42%	46%	10%	1%	0%
	29	3	57%	40%	3%	0%	

The information provided in Table 10 shows a high degree of consistency between the local and audit scoring. Specifically, the percentage of ratings that were exactly the same across local and audit scoring met or exceeded 70% for all items in grades 3 and 5. For grades 4, 6 and 8, the percent perfect agreement was lower, though most agreement was within one score point. A possible explanation for such observation might be because the maximum score points for items in grades 4, 6 and 8 were relatively higher than the maximum score points for items in other grades, under which case agreement is harder to achieve. The percent of scores that differed by two or more points fell below 5% for most items.

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 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	Correlation	Common Variance
3	0.98	0.96
4	0.98	0.96
5	0.99	0.98
6	0.97	0.94
7	0.98	0.96
8	0.97	0.94

The correlations show a very high degree of consistency between the local and audit scoring results. They range from 0.97 to 0.99. Based on these correlations, the amount of common variance between local and audit scoring ranges from 0.94 to 0.98, which means there is little difference between the local and audit scoring results at the total score level. Given that most decisions using test results are based on the total score, these statistics provide valuable evidence of the reliability and consistency in the local and audit scoring results.

Additional Analyses

The results from additional analyses are presented in the appendices. Appendix A contains a detailed item analysis for the ELA 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 ELA audit scores are in Appendix B.

Appendices C, D, and E contain summary item-level information for the ELA assessments. Analyses are computed for all schools and then by scoring model. 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 the state
2. All schools without the New York City schools
3. New York City 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 scoring and the Local scoring for ELA. This information was computed for the various scoring models. The appendices are for:

1. All schools in the state
2. All schools without the New York City schools
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 shows that the samples mirrored the state in these categories. For all grades the obtained proportions in each of the 7 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 range from a low of 0.97 to a high of 0.99. This indicates 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 ELA the largest mean difference between local and audit scoring was 0.5, which occurred in grade 6, item 27. Considering this is a 5-point item, that difference represents only 10% of the maximum points. All but 2 items had mean differences of 0.3 or less.

Appendix C contains the scoring results for each of the scoring models. By inspection it appears that there is little difference between the local and audit scoring results by scoring model. Scoring models 1 and 3 have only one item each where the absolute mean difference is 0.4 or greater. Scoring models 2, 4, and 5 have 4 items each where the absolute mean difference is 0.4 or greater. 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
ELA Item Analysis for Local Scoring

LOCAL SCORING ELA Grade 3 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
21	O	.02	.25	.18	.54	.00	.00	1.28	0.61
26	O	.00	.02	.15	.83	.00	.00	1.80	0.43
27	O	.00	.12	.13	.75	.00	.00	1.63	0.43
28	O	.00	.03	.03	.07	.86	.00	2.75	0.45

LOCAL SCORING ELA Grade 4 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
29	O	.00	.01	.06	.27	.46	.20	2.75	0.69
30	O	.00	.01	.06	.27	.49	.16	2.71	0.71
31	O	.00	.02	.16	.49	.33	.00	2.13	0.63

LOCAL SCORING ELA Grade 5 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
21	O	.01	.15	.25	.59	.00	.00	1.43	0.56
26	O	.00	.12	.45	.44	.00	.00	1.31	0.58
27	O	.00	.27	.26	.30	.16	.00	1.32	0.69

LOCAL SCORING ELA Grade 6 Item Statistics.

Item	Key	B	0	1	2	3	4	5	Mean	PBS
27	O	.00	.00	.03	.12	.33	.37	.14	3.45	0.69
28	O	.01	.01	.05	.15	.34	.32	.12	3.28	0.72
29	O	.00	.01	.13	.49	.37	.00	.00	2.22	0.64

LOCAL SCORING ELA Grade 7 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
3	O	.01	.06	.19	.74	.00	.00	1.68	0.55
22	O	.01	.07	.34	.58	.00	.00	1.52	0.53
33	O	.00	.06	.34	.59	.00	.00	1.54	0.50
34	O	.00	.05	.26	.69	.00	.00	1.65	0.47
35	O	.00	.35	.31	.28	.05	.00	1.02	0.59

LOCAL SCORING ELA Grade 8 Item Statistics.

Item	Key	B	0	1	2	3	4	5	Mean	PBS
27	O	.00	.00	.03	.09	.24	.34	.29	3.75	0.74
28	O	.01	.01	.04	.11	.24	.33	.26	3.62	0.75
29	O	.00	.01	.11	.47	.41	.00	.00	2.26	0.64

Appendix B

ELA Item Analysis for Audit Scoring

Audit Scoring ELA Grade 3 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
21	O	.02	.26	.17	.54	.00	.00	1.27	0.62
26	O	.01	.02	.13	.84	.00	.00	1.81	0.42
27	O	.01	.13	.11	.75	.00	.00	1.61	0.43
28	O	.00	.04	.03	.07	.86	.00	2.75	0.44

Audit Scoring ELA Grade 4 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
29	O	.00	.01	.08	.38	.39	.13	2.53	0.63
30	O	.00	.02	.07	.28	.52	.11	2.61	0.66
31	O	.00	.03	.24	.44	.29	.00	1.99	0.59

Audit Scoring ELA Grade 5 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
21	O	.01	.17	.22	.60	.00	.00	1.43	0.52
26	O	.01	.13	.49	.37	.00	.00	1.22	0.59
27	O	.01	.25	.27	.31	.16	.00	1.35	0.69

Audit Scoring ELA Grade 6 Item Statistics.

Item	Key	B	0	1	2	3	4	5	Mean	PBS
27	O	.01	.01	.04	.21	.44	.25	.05	3.04	0.61
28	O	.01	.01	.08	.22	.40	.21	.08	2.94	0.72
29	O	.00	.02	.22	.51	.25	.00	.00	1.99	0.64

Audit Scoring ELA Grade 7 Item Statistics.

Item	Key	B	0	1	2	3	4	Mean	PBS
3	O	.01	.08	.17	.73	.00	.00	1.63	0.56
22	O	.02	.11	.51	.37	.00	.00	1.24	0.55
33	O	.01	.09	.35	.55	.00	.00	1.46	0.53
34	O	.00	.04	.31	.65	.00	.00	1.61	0.44
35	O	.01	.37	.31	.27	.04	.00	0.97	0.60

Audit Scoring ELA Grade 8 Item Statistics.

Item	Key	B	0	1	2	3	4	5	Mean	PBS
27	O	.00	.00	.02	.08	.27	.37	.24	3.72	0.67
28	O	.01	.01	.06	.13	.32	.33	.14	3.34	0.74
29	O	.00	.01	.16	.45	.37	.00	.00	2.18	0.63

Appendix C

Item Level Statistics for ELA Including All Schools in State

Table C-1. New York State Public Schools Grade 3 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
21	Overall	2	12921	95.2	4.5	99.7	1.3	1.3	0.86	0.87	0.98	0.93
	1	2	6766	94.5	5.1	99.6	1.2	1.1	0.88	0.89	0.98	0.95
	2	2	239	92.1	6.7	98.7	1.3	1.2	0.82	0.83	0.96	0.92
	3	2	4012	95.9	3.8	99.7	1.4	1.4	0.82	0.83	0.98	0.96
	4	2	969	96.6	3.2	99.8	1.3	1.3	0.84	0.84	0.99	0.97
	5	2	935	96.1	3.9	100.0	1.4	1.4	0.84	0.84	0.99	0.96
26	Overall	2	12921	86.1	13.2	99.3	1.8	1.8	0.46	0.46	0.76	0.86
	1	2	6766	85.0	14.6	99.6	1.8	1.8	0.49	0.48	0.79	0.85
	2	2	239	86.6	12.1	98.7	1.8	1.8	0.50	0.48	0.79	0.87
	3	2	4012	87.9	11.4	99.3	1.8	1.9	0.41	0.40	0.72	0.88
	4	2	969	86.1	13.4	99.5	1.9	1.8	0.39	0.46	0.73	0.86
	5	2	935	85.9	11.7	97.5	1.9	1.8	0.39	0.49	0.62	0.86
27	Overall	2	12921	88.6	10.6	99.2	1.6	1.6	0.69	0.71	0.92	0.89
	1	2	6766	88.3	10.8	99.2	1.6	1.6	0.70	0.73	0.93	0.88
	2	2	239	88.3	10.5	98.8	1.5	1.5	0.75	0.79	0.93	0.88
	3	2	4012	89.7	9.7	99.5	1.7	1.7	0.64	0.66	0.92	0.90
	4	2	969	88.2	11.1	99.3	1.6	1.6	0.73	0.73	0.93	0.88
	5	2	935	85.9	12.1	98.0	1.6	1.6	0.68	0.71	0.88	0.86
28	Overall	3	12921	93.6	5.3	98.9	2.8	2.8	0.69	0.70	0.93	0.94
	1	3	6766	92.5	6.4	98.8	2.7	2.7	0.76	0.76	0.94	0.92
	2	3	239	93.3	5.4	98.7	2.7	2.7	0.66	0.72	0.93	0.93
	3	3	4012	95.3	4.0	99.3	2.8	2.8	0.58	0.59	0.93	0.95
	4	3	969	95.3	4.4	99.7	2.8	2.8	0.61	0.60	0.96	0.95
	5	3	935	93.0	3.6	96.7	2.8	2.7	0.62	0.75	0.82	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.

Table C-2. New York State Public Schools Grade 4 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
29	Overall	4	13014	48.7	46.5	95.2	2.8	2.5	0.87	0.87	0.73	0.42
	1	4	7564	50.6	45.0	95.6	2.7	2.5	0.88	0.88	0.75	0.51
	2	4	70	35.7	52.9	88.6	2.7	2.1	0.85	0.54	0.30	0.36
	3	4	3378	46.5	48.7	95.2	2.9	2.6	0.83	0.85	0.70	0.46
	4	4	738	41.7	50.4	92.1	3.0	2.5	0.85	0.77	0.60	0.42
	5	4	1264	47.9	46.5	94.5	2.8	2.5	0.88	0.86	0.71	0.48
30	Overall	4	13014	53.0	43.9	96.9	2.7	2.6	0.85	0.85	0.76	0.53
	1	4	7564	53.5	43.4	96.9	2.6	2.6	0.85	0.87	0.77	0.53
	2	4	70	65.7	32.9	98.6	2.7	2.6	0.73	0.82	0.81	0.66
	3	4	3378	52.3	44.8	97.1	2.9	2.7	0.82	0.80	0.73	0.52
	4	4	738	55.7	39.7	95.4	2.9	2.7	0.81	0.78	0.70	0.56
	5	4	1264	49.6	47.3	96.9	2.8	2.6	0.91	0.87	0.77	0.50
31	Overall	3	13014	53.7	42.9	96.6	2.1	2.0	0.74	0.80	0.69	0.54
	1	3	7564	51.7	44.1	95.7	2.1	1.9	0.74	0.81	0.66	0.52
	2	3	70	58.6	38.6	97.2	2.1	2.2	0.79	0.77	0.75	0.59
	3	3	3378	57.5	40.6	98.1	2.2	2.1	0.72	0.79	0.73	0.58
	4	3	738	56.5	41.6	98.1	2.1	2.1	0.71	0.76	0.71	0.57
	5	3	1264	53.9	42.9	96.8	2.1	2.0	0.73	0.80	0.69	0.54

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.

Table C-3. New York State Public Schools Grade 5 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
21	Overall	2	12536	77.3	21.7	99.0	1.4	1.4	0.75	0.78	0.88	0.61
	1	2	6804	76.0	22.8	98.8	1.4	1.4	0.77	0.79	0.87	0.76
	2	2	416	79.3	20.4	99.8	1.4	1.4	0.76	0.76	0.90	0.79
	3	2	3510	78.8	20.4	99.1	1.5	1.5	0.71	0.74	0.87	0.79
	4	2	716	78.6	20.4	99.0	1.5	1.5	0.71	0.73	0.87	0.79
	5	2	1090	78.5	20.9	99.4	1.4	1.4	0.76	0.79	0.89	0.79
26	Overall	2	12536	76.9	22.2	99.1	1.3	1.2	0.67	0.68	0.84	0.77
	1	2	6804	76.3	22.6	98.8	1.3	1.2	0.68	0.69	0.83	0.76
	2	2	416	80.3	19.2	99.5	1.3	1.3	0.63	0.63	0.85	0.80
	3	2	3510	78.0	21.5	99.5	1.4	1.3	0.67	0.66	0.85	0.78
	4	2	716	78.5	21.1	99.6	1.4	1.4	0.62	0.61	0.82	0.78
	5	2	1090	74.7	24.3	99.0	1.4	1.3	0.66	0.66	0.81	0.75
27	Overall	3	12536	73.4	25.0	98.4	1.3	1.4	1.05	1.04	0.92	0.73
	1	3	6804	73.8	24.6	98.4	1.2	1.2	1.05	1.03	0.92	0.74
	2	3	416	74.5	24.0	98.6	1.3	1.4	0.99	0.98	0.92	0.75
	3	3	3510	73.0	25.0	98.1	1.5	1.5	1.04	1.03	0.92	0.73
	4	3	716	70.7	28.1	98.8	1.6	1.5	0.98	1.01	0.91	0.71
	5	3	1090	73.3	25.6	98.9	1.4	1.5	1.04	1.01	0.92	0.73

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.

Table C-4. New York State Public Schools Grade 6 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
27	Overall	5	11855	38.9	48.5	87.4	3.5	3.0	1.00	0.95	0.65	0.37
	1	5	8115	39.8	48.3	88.1	3.4	3.0	1.01	0.96	0.67	0.40
	2	5	191	22.5	52.9	75.4	3.7	2.8	1.16	0.87	0.54	0.23
	3	5	1181	43.9	47.2	91.1	3.4	3.2	1.00	0.93	0.71	0.44
	4	5	784	32.7	51.3	84.0	3.9	3.2	0.79	0.90	0.43	0.33
	5	5	1584	35.7	48.9	84.6	3.6	3.1	0.97	0.89	0.55	0.36
28	Overall	5	11855	39.9	47.2	87.1	3.3	2.9	1.07	1.08	0.72	0.40
	1	5	8115	41.1	47.2	88.3	3.2	2.9	1.06	1.07	0.73	0.41
	2	5	191	37.2	56.5	93.7	3.1	2.9	1.18	1.03	0.80	0.37
	3	5	1181	39.5	46.7	86.2	3.1	2.8	1.11	1.12	0.73	0.40
	4	5	784	37.1	46.9	84.0	3.7	3.2	1.01	1.00	0.61	0.37
	5	5	1584	35.4	46.6	82.0	3.6	3.0	1.02	1.13	0.64	0.35
29	Overall	3	11855	54.5	42.8	97.4	2.2	2.0	0.70	0.73	0.66	0.55
	1	3	8115	55.6	42.0	97.6	2.2	2.0	0.69	0.73	0.67	0.56
	2	3	191	68.6	30.9	99.5	2.0	2.0	0.72	0.71	0.81	0.69
	3	3	1181	49.7	46.5	96.2	2.2	1.9	0.74	0.79	0.66	0.50
	4	3	784	48.3	47.3	95.6	2.5	2.1	0.61	0.71	0.50	0.48
	5	3	1584	53.8	43.8	97.6	2.3	2.0	0.66	0.72	0.63	0.54

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.

Table C-5. New York State Public Schools Grade 7 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
3	Overall	2	11055	87.6	11.8	99.5	1.7	1.6	0.60	0.65	0.90	0.71
	1	2	5168	86.6	12.8	99.4	1.6	1.6	0.63	0.67	0.90	0.87
	2	2	128	80.5	19.5	100.0	1.6	1.6	0.67	0.68	0.88	0.80
	3	2	929	86.8	12.8	99.6	1.6	1.6	0.64	0.70	0.91	0.87
	4	2	646	87.9	11.6	99.5	1.6	1.6	0.61	0.67	0.91	0.88
	5	2	4184	89.3	10.2	99.5	1.7	1.7	0.54	0.58	0.90	0.89
22	Overall	2	11055	64.7	34.7	99.3	1.5	1.2	0.64	0.66	0.73	0.65
	1	2	5168	64.0	35.2	99.3	1.5	1.2	0.66	0.67	0.74	0.64
	2	2	128	65.6	31.3	96.9	1.6	1.2	0.64	0.75	0.73	0.66
	3	2	929	64.4	34.7	99.0	1.5	1.2	0.67	0.68	0.75	0.64
	4	2	646	62.5	36.5	99.0	1.4	1.1	0.64	0.62	0.69	0.63
	5	2	4184	65.8	33.8	99.6	1.6	1.3	0.60	0.63	0.71	0.66
33	Overall	2	11055	70.0	27.9	97.9	1.5	1.5	0.61	0.67	0.71	0.70
	1	2	5168	67.8	29.4	97.3	1.5	1.4	0.64	0.70	0.71	0.68
	2	2	128	74.2	22.7	96.9	1.5	1.5	0.57	0.70	0.73	0.74
	3	2	929	67.9	30.1	98.1	1.5	1.4	0.63	0.69	0.73	0.68
	4	2	646	71.8	26.0	97.8	1.5	1.4	0.61	0.65	0.72	0.72
	5	2	4184	72.6	26.0	98.6	1.6	1.6	0.57	0.60	0.70	0.73
34	Overall	2	11055	75.4	23.9	99.3	1.7	1.6	0.57	0.57	0.74	0.75
	1	2	5168	74.1	25.0	99.2	1.6	1.6	0.59	0.59	0.74	0.74
	2	2	128	73.4	26.6	100.0	1.7	1.6	0.51	0.56	0.71	0.73
	3	2	929	76.7	22.6	99.4	1.6	1.5	0.58	0.60	0.78	0.77
	4	2	646	72.3	27.6	99.9	1.5	1.5	0.61	0.55	0.74	0.72
	5	2	4184	77.3	22.1	99.5	1.7	1.7	0.54	0.52	0.72	0.77
35	Overall	3	11055	69.4	28.1	97.4	1.0	1.0	0.92	0.90	0.87	0.69
	1	3	5168	68.7	29.0	97.7	0.9	0.8	0.89	0.87	0.86	0.69
	2	3	128	51.6	35.9	87.5	1.4	1.0	0.90	0.91	0.64	0.52
	3	3	929	68.2	28.8	97.1	0.9	0.8	0.91	0.87	0.85	0.68
	4	3	646	68.6	28.0	96.6	1.1	1.0	0.95	0.91	0.86	0.69
	5	3	4184	71.1	26.5	97.6	1.2	1.1	0.91	0.91	0.88	0.71

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.

Table C-6. New York State Public Schools Grade 8 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
27	Overall	5	12245	45.3	46.3	91.6	3.8	3.7	1.10	1.03	0.78	0.45
	1	5	5882	44.9	46.3	91.2	3.6	3.7	1.14	1.04	0.78	0.45
	2	5	63	49.2	49.2	98.4	3.7	3.9	0.81	0.67	0.68	0.49
	3	5	2203	45.3	46.5	91.8	3.7	3.5	1.13	1.10	0.80	0.45
	4	5	759	40.1	47.8	87.9	3.8	3.7	1.07	1.08	0.71	0.40
	5	5	3338	47.2	45.7	92.9	4.1	4.0	0.95	0.91	0.72	0.47
28	Overall	5	12245	42.5	46.8	89.3	3.6	3.3	1.16	1.12	0.78	0.42
	1	5	5882	42.8	46.4	89.2	3.4	3.2	1.21	1.15	0.80	0.43
	2	5	63	58.7	36.5	95.2	3.7	3.6	0.87	0.81	0.76	0.59
	3	5	2203	34.4	48.3	82.7	3.7	3.1	1.18	1.15	0.73	0.34
	4	5	759	48.7	46.0	94.7	3.9	3.7	1.03	1.04	0.80	0.49
	5	5	3338	45.5	46.8	92.3	4.0	3.6	0.98	1.00	0.75	0.46
29	Overall	3	12245	57.4	40.2	97.6	2.3	2.2	0.70	0.74	0.69	0.57
	1	3	5882	58.7	39.2	97.9	2.2	2.1	0.72	0.76	0.72	0.59
	2	3	63	69.8	30.2	100.0	2.5	2.6	0.61	0.61	0.75	0.70
	3	3	2203	52.3	43.7	96.0	2.4	2.1	0.73	0.75	0.64	0.52
	4	3	759	60.5	35.8	96.3	2.2	2.3	0.68	0.71	0.65	0.60
	5	3	3338	57.6	40.6	98.2	2.3	2.3	0.64	0.70	0.64	0.58

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 D

Item Level Statistics for ELA Including All Schools in State Without New York City Schools

**Table D-1. New York State Public Schools (Without NYC) Grade 3 ELA Operational Test 2007:
Inter-rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
21	Overall	2	7746	95.6	4.1	99.7	1.3	1.3	0.84	0.84	0.98	0.93
	1	2	1591	94.6	5.1	99.7	1.2	1.2	0.87	0.87	0.98	0.95
	2	2	239	92.1	6.7	98.8	1.3	1.2	0.82	0.83	0.96	0.92
	3	2	4012	95.9	3.8	99.7	1.4	1.4	0.82	0.83	0.98	0.96
	4	2	969	96.6	3.2	99.8	1.3	1.3	0.84	0.84	0.99	0.97
	5	2	935	96.1	3.9	100.0	1.4	1.4	0.84	0.84	0.99	0.96
26	Overall	2	7746	87.1	12.1	99.2	1.8	1.8	0.42	0.43	0.74	0.87
	1	2	1591	86.5	13.2	99.7	1.8	1.8	0.47	0.46	0.80	0.87
	2	2	239	86.6	12.1	98.7	1.8	1.8	0.50	0.48	0.79	0.87
	3	2	4012	87.9	11.4	99.3	1.8	1.9	0.41	0.40	0.72	0.88
	4	2	969	86.1	13.4	99.5	1.9	1.8	0.39	0.46	0.73	0.86
	5	2	935	85.9	11.7	97.5	1.9	1.8	0.39	0.49	0.62	0.86
27	Overall	2	7746	88.9	10.3	99.2	1.6	1.6	0.68	0.70	0.92	0.89
	1	2	1591	89.3	10.1	99.4	1.6	1.6	0.74	0.75	0.94	0.89
	2	2	239	88.3	10.5	98.8	1.5	1.5	0.75	0.79	0.93	0.88
	3	2	4012	89.7	9.7	99.5	1.7	1.7	0.64	0.66	0.92	0.90
	4	2	969	88.2	11.1	99.3	1.6	1.6	0.73	0.73	0.93	0.88
	5	2	935	85.9	12.1	98.0	1.6	1.6	0.68	0.71	0.88	0.86
28	Overall	3	7746	94.8	4.3	99.1	2.8	2.8	0.62	0.64	0.93	0.95
	1	3	1591	94.3	5.3	99.6	2.7	2.7	0.71	0.71	0.96	0.94
	2	3	239	93.3	5.4	98.7	2.7	2.7	0.66	0.72	0.93	0.93
	3	3	4012	95.3	4.0	99.3	2.8	2.8	0.58	0.59	0.93	0.95
	4	3	969	95.3	4.4	99.7	2.8	2.8	0.61	0.60	0.96	0.95
	5	3	935	93.0	3.6	96.7	2.8	2.7	0.62	0.75	0.82	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.

**Table D-2. New York State Public Schools (Without NYC) Grade 4 ELA Operational Test 2007:
Inter-rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
29	Overall	4	8142	47.6	47.4	95.1	2.9	2.6	0.84	0.84	0.69	0.42
	1	4	2692	50.8	45.4	96.2	2.8	2.6	0.81	0.83	0.71	0.51
	2	4	70	35.7	52.9	88.6	2.7	2.1	0.85	0.54	0.30	0.36
	3	4	3378	46.5	48.7	95.2	2.9	2.6	0.83	0.85	0.70	0.46
	4	4	738	41.7	50.4	92.1	3.0	2.5	0.85	0.77	0.60	0.42
	5	4	1264	47.9	46.5	94.4	2.8	2.5	0.88	0.86	0.71	0.48
30	Overall	4	8142	52.3	44.7	97.0	2.8	2.7	0.84	0.82	0.74	0.52
	1	4	2692	52.4	45.0	97.4	2.8	2.7	0.84	0.82	0.75	0.52
	2	4	70	65.7	32.9	98.6	2.7	2.6	0.73	0.82	0.81	0.66
	3	4	3378	52.3	44.8	97.1	2.9	2.7	0.82	0.80	0.73	0.52
	4	4	738	55.7	39.7	95.4	2.9	2.7	0.81	0.78	0.70	0.56
	5	4	1264	49.6	47.3	96.9	2.8	2.6	0.91	0.87	0.77	0.50
31	Overall	3	8142	55.2	41.9	97.1	2.2	2.0	0.73	0.78	0.70	0.55
	1	3	2692	52.5	43.2	95.7	2.1	2.0	0.74	0.78	0.65	0.52
	2	3	70	58.6	38.6	97.2	2.1	2.2	0.79	0.77	0.75	0.59
	3	3	3378	57.5	40.6	98.1	2.2	2.1	0.72	0.79	0.73	0.58
	4	3	738	56.5	41.6	98.1	2.1	2.1	0.71	0.76	0.71	0.57
	5	3	1264	53.9	42.9	96.8	2.1	2.0	0.73	0.80	0.69	0.54

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.

**Table D-3. New York State Public Schools (Without NYC) Grade 5 ELA Operational Test 2007:
Inter-rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
21	Overall	2	7792	77.7	21.4	99.1	1.5	1.5	0.73	0.74	0.87	0.60
	1	2	2060	74.8	23.9	98.7	1.4	1.5	0.74	0.72	0.84	0.75
	2	2	416	79.3	20.4	99.8	1.4	1.4	0.76	0.76	0.90	0.79
	3	2	3510	78.8	20.4	99.1	1.5	1.5	0.71	0.74	0.87	0.79
	4	2	716	78.6	20.4	99.0	1.5	1.5	0.71	0.73	0.87	0.79
	5	2	1090	78.5	20.9	99.4	1.4	1.4	0.76	0.79	0.89	0.79
26	Overall	2	7792	77.9	21.4	99.4	1.4	1.3	0.65	0.65	0.83	0.78
	1	2	2060	78.9	20.3	99.3	1.4	1.3	0.63	0.62	0.83	0.79
	2	2	416	80.3	19.2	99.5	1.3	1.3	0.63	0.63	0.85	0.80
	3	2	3510	78.0	21.5	99.5	1.4	1.3	0.67	0.66	0.85	0.78
	4	2	716	78.5	21.1	99.6	1.4	1.4	0.62	0.61	0.82	0.78
	5	2	1090	74.7	24.3	99.0	1.4	1.3	0.66	0.66	0.81	0.75
27	Overall	3	7792	73.4	25.1	98.4	1.5	1.5	1.03	1.01	0.92	0.73
	1	3	2060	74.7	24.0	98.7	1.6	1.6	1.03	0.99	0.92	0.75
	2	3	416	74.5	24.0	98.6	1.3	1.4	0.99	0.98	0.92	0.75
	3	3	3510	73.0	25.0	98.1	1.5	1.5	1.04	1.03	0.92	0.73
	4	3	716	70.7	28.1	98.8	1.6	1.5	0.98	1.01	0.91	0.71
	5	3	1090	73.3	25.6	98.9	1.4	1.5	1.04	1.01	0.92	0.73

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.

**Table D-4. New York State Public Schools (Without NYC) Grade 6 ELA Operational Test 2007:
Inter-rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
27	Overall	5	7474	37.6	49.0	86.6	3.5	3.1	0.99	0.93	0.62	0.36
	1	5	3734	38.3	49.0	87.3	3.5	3.1	1.01	0.96	0.64	0.38
	2	5	191	22.5	52.9	75.4	3.7	2.8	1.16	0.87	0.54	0.23
	3	5	1181	43.9	47.2	91.1	3.4	3.2	1.00	0.93	0.71	0.44
	4	5	784	32.7	51.3	84.0	3.9	3.2	0.79	0.90	0.43	0.33
	5	5	1584	35.7	48.9	84.6	3.6	3.1	0.97	0.89	0.55	0.36
28	Overall	5	7474	39.9	47.1	87.0	3.4	3.0	1.07	1.06	0.71	0.40
	1	5	3734	42.7	47.0	89.7	3.3	3.0	1.05	1.02	0.74	0.43
	2	5	191	37.2	56.5	93.7	3.1	2.9	1.18	1.03	0.80	0.37
	3	5	1181	39.5	46.7	86.2	3.1	2.8	1.11	1.12	0.73	0.40
	4	5	784	37.1	46.9	84.0	3.7	3.2	1.01	1.00	0.61	0.37
	5	5	1584	35.4	46.6	82.0	3.6	3.0	1.02	1.13	0.64	0.35
29	Overall	3	7474	54.2	42.9	97.1	2.3	2.0	0.70	0.73	0.65	0.54
	1	3	3734	56.2	41.2	97.4	2.2	2.0	0.69	0.71	0.65	0.56
	2	3	191	68.6	30.9	99.5	2.0	2.0	0.72	0.71	0.81	0.69
	3	3	1181	49.7	46.5	96.2	2.2	1.9	0.74	0.79	0.66	0.50
	4	3	784	48.3	47.3	95.6	2.5	2.1	0.61	0.71	0.50	0.48
	5	3	1584	53.8	43.8	97.6	2.3	2.0	0.66	0.72	0.63	0.54

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.

**Table D-5. New York State Public Schools (Without NYC) Grade 7 ELA Operational Test 2007:
Inter-rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
3	Overall	2	6908	88.5	11.0	99.5	1.7	1.7	0.57	0.62	0.90	0.71
	1	2	1021	87.9	11.6	99.4	1.7	1.7	0.57	0.60	0.89	0.88
	2	2	128	80.5	19.5	100.0	1.6	1.6	0.67	0.68	0.88	0.80
	3	2	929	86.8	12.8	99.6	1.6	1.6	0.64	0.70	0.91	0.87
	4	2	646	87.9	11.6	99.5	1.6	1.6	0.61	0.67	0.91	0.88
	5	2	4184	89.3	10.2	99.5	1.7	1.7	0.54	0.58	0.90	0.89
22	Overall	2	6908	65.1	34.3	99.4	1.5	1.3	0.62	0.65	0.72	0.65
	1	2	1021	64.6	34.8	99.4	1.5	1.3	0.61	0.64	0.71	0.65
	2	2	128	65.6	31.3	96.9	1.6	1.2	0.64	0.75	0.73	0.66
	3	2	929	64.4	34.7	99.0	1.5	1.2	0.67	0.68	0.75	0.64
	4	2	646	62.5	36.5	99.0	1.4	1.1	0.64	0.62	0.69	0.63
	5	2	4184	65.8	33.8	99.6	1.6	1.3	0.60	0.63	0.71	0.66
33	Overall	2	6908	72.0	26.3	98.3	1.6	1.5	0.58	0.63	0.71	0.72
	1	2	1021	73.1	25.0	98.0	1.6	1.6	0.55	0.62	0.69	0.73
	2	2	128	74.2	22.7	96.9	1.5	1.5	0.57	0.70	0.73	0.74
	3	2	929	67.9	30.1	98.1	1.5	1.4	0.63	0.69	0.73	0.68
	4	2	646	71.8	26.0	97.8	1.5	1.4	0.61	0.65	0.72	0.72
	5	2	4184	72.6	26.0	98.6	1.6	1.6	0.57	0.60	0.70	0.73
34	Overall	2	6908	76.2	23.2	99.5	1.7	1.6	0.55	0.54	0.73	0.76
	1	2	1021	74.2	25.1	99.3	1.7	1.7	0.55	0.53	0.69	0.74
	2	2	128	73.4	26.6	100.0	1.7	1.6	0.51	0.56	0.71	0.73
	3	2	929	76.7	22.6	99.4	1.6	1.5	0.58	0.60	0.78	0.77
	4	2	646	72.3	27.6	99.9	1.5	1.5	0.61	0.55	0.74	0.72
	5	2	4184	77.3	22.1	99.5	1.7	1.7	0.54	0.52	0.72	0.77
35	Overall	3	6908	69.5	27.6	97.1	1.1	1.1	0.93	0.92	0.87	0.70
	1	3	1021	67.1	29.7	96.8	1.2	1.2	0.97	0.94	0.87	0.67
	2	3	128	51.6	35.9	87.5	1.4	1.0	0.90	0.91	0.64	0.52
	3	3	929	68.2	28.8	97.1	0.9	0.8	0.91	0.87	0.85	0.68
	4	3	646	68.6	28.0	96.6	1.1	1.0	0.95	0.91	0.86	0.69
	5	3	4184	71.1	26.5	97.6	1.2	1.1	0.91	0.91	0.88	0.71

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.

**Table D-6. New York State Public Schools (Without NYC) Grade 8 ELA Operational Test 2007:
Inter-rater Agreement**

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
27	Overall	5	8593	45.1	46.2	91.3	3.9	3.8	1.05	1.01	0.75	0.45
	1	5	2230	43.5	45.8	89.3	3.7	3.9	1.05	0.96	0.71	0.43
	2	5	63	49.2	49.2	98.4	3.7	3.9	0.81	0.67	0.68	0.49
	3	5	2203	45.3	46.5	91.8	3.7	3.5	1.13	1.10	0.80	0.45
	4	5	759	40.1	47.8	87.9	3.8	3.7	1.07	1.08	0.71	0.40
	5	5	3338	47.2	45.7	92.9	4.1	4.0	0.95	0.91	0.72	0.47
28	Overall	5	8593	43.1	46.9	90.1	3.8	3.5	1.07	1.07	0.76	0.43
	1	5	2230	45.8	46.4	92.2	3.6	3.5	1.08	1.04	0.79	0.46
	2	5	63	58.7	36.5	95.2	3.7	3.6	0.87	0.81	0.76	0.59
	3	5	2203	34.4	48.3	82.7	3.7	3.1	1.18	1.15	0.73	0.34
	4	5	759	48.7	46.0	94.7	3.9	3.7	1.03	1.04	0.80	0.49
	5	5	3338	45.5	46.8	92.3	4.0	3.6	0.98	1.00	0.75	0.46
29	Overall	3	8593	57.0	40.5	97.6	2.3	2.2	0.68	0.72	0.65	0.57
	1	3	2230	59.4	39.2	98.6	2.3	2.3	0.68	0.69	0.68	0.59
	2	3	63	69.8	30.2	100.0	2.5	2.6	0.61	0.61	0.75	0.70
	3	3	2203	52.3	43.7	96.0	2.4	2.1	0.73	0.75	0.64	0.52
	4	3	759	60.5	35.8	96.3	2.2	2.3	0.68	0.71	0.65	0.60
	5	3	3338	57.6	40.6	98.2	2.3	2.3	0.64	0.70	0.64	0.58

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 E

**Item Level Statistics for ELA Including
New York City Schools Only**

Table E-1. NYC Public Schools Grades 3 - 8 ELA Operational Test 2007: Inter-rater Agreement

Item #	Scoring Model	Score Points	Total N	Agreement (%)			RS Mean		RS SD		Intra-Class Correlation	Weighted Kappa
				Exact	Approx.	Total	Local	Audit	Local	Audit		
Grade 3												
21	NYC	2	5175	94.5	5.2	99.6	1.2	1.1	0.88	0.90	0.98	0.94
26	NYC	2	5175	84.5	15.0	99.5	1.8	1.8	0.50	0.49	0.79	0.85
27	NYC	2	5175	88.1	11.1	99.1	1.6	1.6	0.69	0.73	0.92	0.88
28	NYC	3	5175	91.9	6.7	98.6	2.7	2.7	0.77	0.77	0.94	0.92
Grade 4												
29	NYC	4	4872	50.6	44.8	95.4	2.6	2.4	0.90	0.91	0.76	0.51
30	NYC	4	4872	54.0	42.6	96.6	2.6	2.5	0.85	0.89	0.77	0.54
31	NYC	3	4872	51.2	44.5	95.7	2.1	1.9	0.74	0.82	0.66	0.51
Grade 5												
21	NYC	2	4744	76.6	22.3	98.8	1.4	1.3	0.78	0.82	0.88	0.77
26	NYC	2	4744	75.1	23.5	98.7	1.2	1.1	0.70	0.70	0.83	0.75
27	NYC	3	4744	73.4	24.9	98.3	1.1	1.1	1.03	1.01	0.92	0.73
Grade 6												
27	NYC	5	4381	41.2	47.7	88.8	3.3	2.9	1.00	0.96	0.68	0.41
28	NYC	5	4381	39.8	47.3	87.1	3.2	2.8	1.07	1.09	0.72	0.40
29	NYC	3	4381	55.1	42.7	97.8	2.1	1.9	0.69	0.75	0.67	0.55
Grade 7												
3	NYC	2	4147	86.3	13.1	99.4	1.6	1.6	0.64	0.69	0.90	0.86
22	NYC	2	4147	63.9	35.4	99.3	1.4	1.2	0.68	0.68	0.75	0.64
33	NYC	2	4147	66.6	30.5	97.1	1.5	1.3	0.65	0.71	0.71	0.67
34	NYC	2	4147	74.1	25.0	99.1	1.6	1.6	0.59	0.60	0.75	0.74
35	NYC	3	4147	69.0	28.9	97.9	0.8	0.8	0.85	0.84	0.85	0.69
Grade 8												
27	NYC	5	3652	45.8	46.5	92.3	3.5	3.5	1.19	1.06	0.81	0.46
28	NYC	5	3652	40.9	46.5	87.4	3.3	3.1	1.27	1.19	0.79	0.41
29	NYC	3	3652	58.3	39.3	97.6	2.2	2.0	0.74	0.79	0.73	0.58

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 ELA Including All Schools in State

**Table F-1. New York State Public Schools Grade 3 ELA Operational Test 2005: Proportions of Score Differences
[Audit Scoring minus Local Scoring]**

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
21	Overall			0.00	0.00	0.03	0.95	0.02	0.00	0.00		
	1			0.00	0.00	0.03	0.95	0.02	0.00	0.00		
	2			0.00	0.01	0.04	0.92	0.03	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	0.97	0.02	0.00	0.00		
	5			0.00	0.00	0.02	0.96	0.02	0.00	0.00		
26	Overall			0.00	0.00	0.06	0.86	0.07	0.00	0.00		
	1			0.00	0.00	0.06	0.85	0.09	0.00	0.00		
	2			0.00	0.00	0.04	0.87	0.08	0.01	0.00		
	3			0.00	0.00	0.05	0.88	0.07	0.00	0.00		
	4			0.00	0.01	0.09	0.86	0.05	0.00	0.00		
	5			0.00	0.02	0.06	0.86	0.05	0.00	0.00		
27	Overall			0.00	0.01	0.05	0.89	0.05	0.00	0.00		
	1			0.00	0.01	0.06	0.88	0.05	0.00	0.00		
	2			0.00	0.01	0.05	0.88	0.05	0.00	0.00		
	3			0.00	0.00	0.05	0.90	0.05	0.00	0.00		
	4			0.00	0.00	0.06	0.88	0.05	0.00	0.00		
	5			0.00	0.02	0.04	0.86	0.08	0.00	0.00		
28	Overall			0.00	0.00	0.02	0.94	0.03	0.00	0.00		
	1			0.00	0.00	0.03	0.92	0.04	0.00	0.00		
	2			0.00	0.00	0.03	0.93	0.02	0.00	0.00		
	3			0.00	0.00	0.02	0.95	0.02	0.00	0.00		
	4			0.00	0.00	0.02	0.95	0.02	0.00	0.00		
	5			0.02	0.01	0.02	0.93	0.02	0.00	0.00		

Table F-2. New York State Public Schools Grade 4 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
29	Overall		0.00	0.00	0.04	0.32	0.49	0.15	0.01	0.00	0.00	
	1		0.00	0.00	0.03	0.29	0.51	0.16	0.01	0.00	0.00	
	2		0.00	0.00	0.10	0.47	0.36	0.06	0.01	0.00	0.00	
	3		0.00	0.00	0.04	0.36	0.46	0.13	0.01	0.00	0.00	
	4		0.00	0.00	0.08	0.42	0.42	0.09	0.00	0.00	0.00	
	5		0.00	0.00	0.05	0.33	0.48	0.13	0.00	0.00	0.00	
30	Overall		0.00	0.00	0.02	0.26	0.53	0.18	0.01	0.00	0.00	
	1		0.00	0.00	0.02	0.24	0.53	0.20	0.01	0.00	0.00	
	2		0.00	0.00	0.00	0.21	0.66	0.11	0.01	0.00	0.00	
	3		0.00	0.00	0.02	0.30	0.52	0.15	0.01	0.00	0.00	
	4		0.00	0.00	0.04	0.25	0.56	0.15	0.01	0.00	0.00	
	5		0.00	0.00	0.02	0.31	0.50	0.16	0.01	0.00	0.00	
31	Overall		0.00	0.00	0.02	0.27	0.54	0.16	0.01	0.00	0.00	
	1		0.00	0.00	0.03	0.27	0.52	0.17	0.01	0.00	0.00	
	2		0.00	0.00	0.00	0.20	0.59	0.19	0.03	0.00	0.00	
	3		0.00	0.00	0.02	0.27	0.58	0.14	0.00	0.00	0.00	
	4		0.00	0.00	0.01	0.25	0.57	0.17	0.01	0.00	0.00	
	5		0.00	0.00	0.02	0.25	0.54	0.17	0.01	0.00	0.00	

**Table F-3. New York State Public Schools Grade 5 ELA Operational Test 2007: Proportions of Score Differences
[Audit Scoring minus Local Scoring]**

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
21	Overall			0.00	0.01	0.11	0.77	0.11	0.00	0.00		
	1			0.00	0.01	0.11	0.76	0.12	0.01	0.00		
	2			0.00	0.00	0.11	0.79	0.09	0.00	0.00		
	3			0.00	0.01	0.11	0.79	0.09	0.00	0.00		
	4			0.00	0.01	0.09	0.79	0.11	0.00	0.00		
	5			0.00	0.00	0.10	0.79	0.11	0.00	0.00		
26	Overall			0.00	0.01	0.15	0.77	0.08	0.00	0.00		
	1			0.00	0.01	0.16	0.76	0.07	0.00	0.00		
	2			0.00	0.00	0.12	0.80	0.08	0.00	0.00		
	3			0.00	0.00	0.13	0.78	0.08	0.00	0.00		
	4			0.00	0.00	0.12	0.78	0.09	0.00	0.00		
	5			0.00	0.01	0.17	0.75	0.08	0.00	0.00		
27	Overall			0.00	0.01	0.12	0.73	0.13	0.01	0.00		
	1			0.00	0.01	0.13	0.74	0.12	0.01	0.00		
	2			0.00	0.00	0.07	0.75	0.17	0.01	0.00		
	3			0.00	0.01	0.11	0.73	0.14	0.01	0.00		
	4			0.00	0.01	0.18	0.71	0.10	0.00	0.00		
	5			0.00	0.00	0.09	0.73	0.16	0.01	0.00		

Table F-4. New York State Public Schools Grade 6 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
27	Overall	0.00	0.00	0.01	0.10	0.35	0.39	0.13	0.02	0.00	0.00	0.00
	1	0.00	0.00	0.01	0.09	0.35	0.40	0.13	0.02	0.00	0.00	0.00
	2	0.00	0.00	0.02	0.23	0.46	0.23	0.07	0.01	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.06	0.29	0.44	0.18	0.02	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.15	0.44	0.33	0.07	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.01	0.12	0.37	0.36	0.11	0.02	0.00	0.00	0.00
28	Overall	0.00	0.00	0.01	0.10	0.32	0.40	0.15	0.02	0.00	0.00	0.00
	1	0.00	0.00	0.01	0.09	0.32	0.41	0.15	0.02	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.05	0.36	0.37	0.21	0.02	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.09	0.29	0.40	0.18	0.03	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.12	0.35	0.37	0.12	0.02	0.00	0.00	0.00
	5	0.00	0.00	0.02	0.15	0.36	0.35	0.11	0.01	0.00	0.00	0.00
29	Overall	0.00	0.00	0.00	0.02	0.31	0.55	0.12	0.00	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.02	0.29	0.56	0.13	0.00	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.16	0.69	0.15	0.01	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.03	0.34	0.50	0.12	0.01	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.04	0.40	0.48	0.07	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.02	0.34	0.54	0.10	0.00	0.00	0.00	0.00

Table F-5. New York State Public Schools Grade 7 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
3	Overall			0.00	0.00	0.07	0.88	0.05	0.00	0.00		
	1			0.00	0.00	0.08	0.87	0.05	0.00	0.00		
	2			0.00	0.00	0.13	0.80	0.07	0.00	0.00		
	3			0.00	0.00	0.09	0.87	0.03	0.00	0.00		
	4			0.00	0.00	0.07	0.88	0.04	0.00	0.00		
22	Overall			0.00	0.01	0.30	0.65	0.05	0.00	0.00		
	1			0.00	0.01	0.30	0.64	0.05	0.00	0.00		
	2			0.00	0.03	0.28	0.66	0.03	0.00	0.00		
	3			0.00	0.01	0.31	0.64	0.04	0.00	0.00		
	4			0.00	0.01	0.34	0.63	0.03	0.00	0.00		
33	Overall			0.00	0.02	0.17	0.70	0.11	0.00	0.00		
	1			0.00	0.02	0.19	0.68	0.11	0.01	0.00		
	2			0.00	0.03	0.11	0.74	0.12	0.00	0.00		
	3			0.00	0.02	0.20	0.68	0.10	0.00	0.00		
	4			0.00	0.01	0.18	0.72	0.08	0.01	0.00		
34	Overall			0.00	0.00	0.14	0.75	0.10	0.00	0.00		
	1			0.00	0.00	0.15	0.74	0.10	0.00	0.00		
	2			0.00	0.00	0.18	0.73	0.09	0.00	0.00		
	3			0.00	0.00	0.13	0.77	0.09	0.00	0.00		
	4			0.00	0.00	0.13	0.72	0.14	0.00	0.00		
35	Overall			0.00	0.01	0.16	0.69	0.12	0.01	0.00		
	1			0.00	0.01	0.17	0.69	0.12	0.01	0.00		
	2			0.01	0.11	0.23	0.52	0.13	0.01	0.00		
	3			0.00	0.01	0.19	0.68	0.10	0.01	0.00		
	4			0.00	0.02	0.16	0.69	0.12	0.01	0.00		

**Table F-6. New York State Public Schools Grade 8 ELA Operational Test 2007: Proportions of Score Differences
[Audit Scoring minus Local Scoring]**

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
27	Overall	0.00	0.00	0.00	0.04	0.24	0.45	0.22	0.04	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.03	0.21	0.45	0.26	0.05	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.14	0.49	0.35	0.02	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.05	0.31	0.45	0.16	0.02	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.06	0.27	0.40	0.21	0.05	0.01	0.00	0.00
	5	0.00	0.00	0.00	0.04	0.25	0.47	0.21	0.03	0.00	0.00	0.00
28	Overall	0.00	0.00	0.01	0.08	0.31	0.42	0.15	0.02	0.00	0.00	0.00
	1	0.00	0.00	0.01	0.07	0.27	0.43	0.19	0.03	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.05	0.19	0.59	0.17	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.15	0.39	0.34	0.09	0.01	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.04	0.30	0.49	0.16	0.01	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.06	0.34	0.46	0.13	0.01	0.00	0.00	0.00
29	Overall	0.00	0.00	0.00	0.02	0.23	0.57	0.17	0.01	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.01	0.22	0.59	0.17	0.01	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.13	0.70	0.17	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.03	0.33	0.52	0.11	0.01	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.02	0.16	0.60	0.19	0.01	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.01	0.21	0.58	0.20	0.01	0.00	0.00	0.00

Appendix G

Item Level Differences for ELA All Schools in State Without New York City Schools

**Table G-1. New York State Public Schools Grade 3 ELA Operational Test 2007: Proportions of Score Differences
[Audit Scoring minus Local Scoring]**

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
21	Overall			0.00	0.00	0.02	0.96	0.02	0.00	0.00		
	1			0.00	0.00	0.03	0.95	0.03	0.00	0.00		
	2			0.00	0.01	0.04	0.92	0.03	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	0.97	0.02	0.00	0.00		
	5			0.00	0.00	0.02	0.96	0.02	0.00	0.00		
26	Overall			0.00	0.01	0.06	0.87	0.06	0.00	0.00		
	1			0.00	0.00	0.05	0.87	0.08	0.00	0.00		
	2			0.00	0.00	0.04	0.87	0.08	0.01	0.00		
	3			0.00	0.00	0.05	0.88	0.07	0.00	0.00		
	4			0.00	0.01	0.09	0.86	0.05	0.00	0.00		
	5			0.00	0.02	0.06	0.86	0.05	0.00	0.00		
27	Overall			0.00	0.01	0.05	0.89	0.05	0.00	0.00		
	1			0.00	0.01	0.05	0.89	0.05	0.00	0.00		
	2			0.00	0.01	0.05	0.88	0.05	0.00	0.00		
	3			0.00	0.00	0.05	0.90	0.05	0.00	0.00		
	4			0.00	0.00	0.06	0.88	0.05	0.00	0.00		
	5			0.00	0.02	0.04	0.86	0.08	0.00	0.00		
28	Overall			0.00	0.00	0.02	0.95	0.02	0.00	0.00		
	1			0.00	0.00	0.02	0.94	0.03	0.00	0.00		
	2			0.00	0.00	0.03	0.93	0.02	0.00	0.00		
	3			0.00	0.00	0.02	0.95	0.02	0.00	0.00		
	4			0.00	0.00	0.02	0.95	0.02	0.00	0.00		
	5			0.02	0.01	0.02	0.93	0.02	0.00	0.00		

Table G-2. New York State Public Schools Grade 4 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
29	Overall		0.00	0.00	0.04	0.34	0.48	0.14	0.01	0.00	0.00	
	1		0.00	0.00	0.03	0.29	0.51	0.17	0.01	0.00	0.00	
	2		0.00	0.00	0.10	0.47	0.36	0.06	0.01	0.00	0.00	
	3		0.00	0.00	0.04	0.36	0.46	0.13	0.01	0.00	0.00	
	4		0.00	0.00	0.08	0.42	0.42	0.09	0.00	0.00	0.00	
	5		0.00	0.00	0.05	0.33	0.48	0.13	0.00	0.00	0.00	
30	Overall		0.00	0.00	0.02	0.28	0.52	0.17	0.01	0.00	0.00	
	1		0.00	0.00	0.02	0.25	0.52	0.20	0.01	0.00	0.00	
	2		0.00	0.00	0.00	0.21	0.66	0.11	0.01	0.00	0.00	
	3		0.00	0.00	0.02	0.30	0.52	0.15	0.01	0.00	0.00	
	4		0.00	0.00	0.04	0.25	0.56	0.15	0.01	0.00	0.00	
	5		0.00	0.00	0.02	0.31	0.50	0.16	0.01	0.00	0.00	
31	Overall		0.00	0.00	0.02	0.26	0.55	0.16	0.01	0.00	0.00	
	1		0.00	0.00	0.03	0.27	0.52	0.17	0.01	0.00	0.00	
	2		0.00	0.00	0.00	0.20	0.59	0.19	0.03	0.00	0.00	
	3		0.00	0.00	0.02	0.27	0.58	0.14	0.00	0.00	0.00	
	4		0.00	0.00	0.01	0.25	0.57	0.17	0.01	0.00	0.00	
	5		0.00	0.00	0.02	0.25	0.54	0.17	0.01	0.00	0.00	

Table G-3. New York State Public Schools Grade 5 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
21	Overall			0.00	0.00	0.10	0.78	0.11	0.00	0.00		
	1			0.00	0.00	0.09	0.75	0.15	0.01	0.00		
	2			0.00	0.00	0.11	0.79	0.09	0.00	0.00		
	3			0.00	0.01	0.11	0.79	0.09	0.00	0.00		
	4			0.00	0.01	0.09	0.79	0.11	0.00	0.00		
	5			0.00	0.00	0.10	0.79	0.11	0.00	0.00		
26	Overall			0.00	0.00	0.13	0.78	0.08	0.00	0.00		
	1			0.00	0.00	0.13	0.79	0.08	0.00	0.00		
	2			0.00	0.00	0.12	0.80	0.08	0.00	0.00		
	3			0.00	0.00	0.13	0.78	0.08	0.00	0.00		
	4			0.00	0.00	0.12	0.78	0.09	0.00	0.00		
	5			0.00	0.01	0.17	0.75	0.08	0.00	0.00		
27	Overall			0.00	0.01	0.11	0.73	0.14	0.01	0.00		
	1			0.00	0.00	0.12	0.75	0.12	0.01	0.00		
	2			0.00	0.00	0.07	0.75	0.17	0.01	0.00		
	3			0.00	0.01	0.11	0.73	0.14	0.01	0.00		
	4			0.00	0.01	0.18	0.71	0.10	0.00	0.00		
	5			0.00	0.00	0.09	0.73	0.16	0.01	0.00		

Table G-4. New York State Public Schools Grade 6 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
27	Overall	0.00	0.00	0.01	0.11	0.36	0.38	0.13	0.02	0.00	0.00	0.00
	1	0.00	0.00	0.01	0.10	0.35	0.38	0.14	0.02	0.00	0.00	0.00
	2	0.00	0.00	0.02	0.23	0.46	0.23	0.07	0.01	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.06	0.29	0.44	0.18	0.02	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.15	0.44	0.33	0.07	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.01	0.12	0.37	0.36	0.11	0.02	0.00	0.00	0.00
28	Overall	0.00	0.00	0.01	0.10	0.32	0.40	0.15	0.02	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.07	0.30	0.43	0.17	0.02	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.05	0.36	0.37	0.21	0.02	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.09	0.29	0.40	0.18	0.03	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.12	0.35	0.37	0.12	0.02	0.00	0.00	0.00
	5	0.00	0.00	0.02	0.15	0.36	0.35	0.11	0.01	0.00	0.00	0.00
29	Overall	0.00	0.00	0.00	0.03	0.31	0.54	0.11	0.00	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.02	0.28	0.56	0.13	0.00	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.16	0.69	0.15	0.01	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.03	0.34	0.50	0.12	0.01	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.04	0.40	0.48	0.07	0.00	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.02	0.34	0.54	0.10	0.00	0.00	0.00	0.00

Table G-5. New York State Public Schools Grade 7 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
3	Overall			0.00	0.00	0.06	0.88	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	0.13	0.80	0.07	0.00	0.00		
	3			0.00	0.00	0.09	0.87	0.03	0.00	0.00		
	4			0.00	0.00	0.07	0.88	0.04	0.00	0.00		
	5			0.00	0.00	0.06	0.89	0.05	0.00	0.00		
22	Overall			0.00	0.01	0.30	0.65	0.04	0.00	0.00		
	1			0.00	0.00	0.30	0.65	0.05	0.00	0.00		
	2			0.00	0.03	0.28	0.66	0.03	0.00	0.00		
	3			0.00	0.01	0.31	0.64	0.04	0.00	0.00		
	4			0.00	0.01	0.34	0.63	0.03	0.00	0.00		
	5			0.00	0.00	0.29	0.66	0.04	0.00	0.00		
33	Overall			0.00	0.01	0.15	0.72	0.12	0.01	0.00		
	1			0.00	0.01	0.13	0.73	0.12	0.01	0.00		
	2			0.00	0.03	0.11	0.74	0.12	0.00	0.00		
	3			0.00	0.02	0.20	0.68	0.10	0.00	0.00		
	4			0.00	0.01	0.18	0.72	0.08	0.01	0.00		
	5			0.00	0.01	0.14	0.73	0.12	0.00	0.00		
34	Overall			0.00	0.00	0.13	0.76	0.11	0.00	0.00		
	1			0.00	0.00	0.11	0.74	0.14	0.00	0.00		
	2			0.00	0.00	0.18	0.73	0.09	0.00	0.00		
	3			0.00	0.00	0.13	0.77	0.09	0.00	0.00		
	4			0.00	0.00	0.13	0.72	0.14	0.00	0.00		
	5			0.00	0.00	0.13	0.77	0.10	0.00	0.00		
35	Overall			0.00	0.02	0.16	0.70	0.12	0.01	0.00		
	1			0.00	0.02	0.18	0.67	0.11	0.01	0.00		
	2			0.01	0.11	0.23	0.52	0.13	0.01	0.00		
	3			0.00	0.01	0.19	0.68	0.10	0.01	0.00		
	4			0.00	0.02	0.16	0.69	0.12	0.01	0.00		
	5			0.00	0.01	0.15	0.71	0.12	0.01	0.00		

**Table G-6. New York State Public Schools Grade 8 ELA Operational Test 2007: Proportions of Score Differences
[Audit Scoring minus Local Scoring]**

Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
27	Overall	0.00	0.00	0.00	0.04	0.25	0.45	0.22	0.04	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.03	0.17	0.43	0.29	0.07	0.01	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.14	0.49	0.35	0.02	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.05	0.31	0.45	0.16	0.02	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.06	0.27	0.40	0.21	0.05	0.01	0.00	0.00
	5	0.00	0.00	0.00	0.04	0.25	0.47	0.21	0.03	0.00	0.00	0.00
28	Overall	0.00	0.00	0.01	0.08	0.33	0.43	0.14	0.01	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.05	0.28	0.46	0.19	0.03	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.05	0.19	0.59	0.17	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.01	0.15	0.39	0.34	0.09	0.01	0.00	0.00	0.00
	4	0.00	0.00	0.01	0.04	0.30	0.49	0.16	0.01	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.06	0.34	0.46	0.13	0.01	0.00	0.00	0.00
29	Overall	0.00	0.00	0.00	0.02	0.23	0.57	0.18	0.01	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.01	0.19	0.59	0.20	0.01	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00	0.13	0.70	0.17	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.03	0.33	0.52	0.11	0.01	0.00	0.00	0.00
	4	0.00	0.00	0.00	0.02	0.16	0.60	0.19	0.01	0.00	0.00	0.00
	5	0.00	0.00	0.00	0.01	0.21	0.58	0.20	0.01	0.00	0.00	0.00

Appendix H

Item Level Differences for ELA Including New York City Schools Only

Table H-1. New York State Public Schools Grade 3 – 8 ELA Operational Test 2007: Proportions of Score Differences [Audit Scoring minus Local Scoring]

Grade 3												
Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
21	NYC			0.00	0.00	0.03	0.94	0.02	0.00	0.00		
26	NYC			0.00	0.00	0.06	0.85	0.09	0.00	0.00		
27	NYC			0.00	0.01	0.06	0.88	0.05	0.00	0.00		
28	NYC			0.00	0.00	0.03	0.92	0.04	0.00	0.00		
Grade 4												
Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
29	NYC		0.00	0.00	0.03	0.28	0.51	0.16	0.01	0.00	0.00	
30	NYC		0.00	0.00	0.02	0.23	0.54	0.19	0.01	0.00	0.00	
31	NYC		0.00	0.00	0.03	0.28	0.51	0.17	0.01	0.00	0.00	
Grade 5												
Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
21	NYC			0.00	0.01	0.12	0.77	0.10	0.00	0.00		
26	NYC			0.00	0.01	0.17	0.75	0.07	0.00	0.00		
27	NYC			0.00	0.01	0.13	0.73	0.11	0.01	0.00		
Grade 6												
Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
27	NYC	0.00	0.00	0.01	0.09	0.35	0.41	0.13	0.01	0.00	0.00	0.00
28	NYC	0.00	0.00	0.01	0.09	0.33	0.40	0.14	0.02	0.00	0.00	0.00
29	NYC	0.00	0.00	0.00	0.02	0.30	0.55	0.13	0.00	0.00	0.00	0.00
Grade 7												
Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
3	NYC			0.00	0.00	0.08	0.86	0.05	0.00	0.00		
22	NYC			0.00	0.01	0.30	0.64	0.06	0.00	0.00		
33	NYC			0.00	0.02	0.20	0.67	0.11	0.00	0.00		
34	NYC			0.00	0.00	0.16	0.74	0.09	0.00	0.00		
35	NYC			0.00	0.01	0.16	0.69	0.12	0.01	0.00		
Grade 8												
Item #	Scoring Model	-5	-4	-3	-2	-1	0	1	2	3	4	5
27	NYC	0.00	0.00	0.00	0.03	0.23	0.46	0.24	0.04	0.00	0.00	0.00
28	NYC	0.00	0.00	0.01	0.08	0.27	0.41	0.20	0.04	0.00	0.00	0.00
29	NYC	0.00	0.00	0.00	0.02	0.24	0.58	0.15	0.01	0.00	0.00	0.00