

Appendix D: Rater Reliability and Severity

The CSAP test design framework includes a variety of different item types, including short response and extended constructed response items. Although constructed-response items greatly enhance the construct and instructional validity of CSAP, reliability of hand-scoring items should be closely examined and documented. Through the ongoing process of training and research analyses, evidence of reliability of hand-scoring is continuously gathered. Many training and monitoring techniques are used to ensure scoring reliability and accuracy. Scoring guides are carefully developed and refined; scorers are trained, calibrated, and monitored throughout the scoring process; and rater reliability indices are generated and examined. Reliability for constructed responses items are typically examined by calculating indices of inter-rater agreement: the reliability with which human raters assign scores to student responses. For this analysis, a certain percentage of student responses are scored by two raters.

Inter-rater Reliability

Evidence supporting inter-rater reliability of CSAP assessments is presented in terms of raw score means, raw score standard deviations, and percentages of exact and adjacent agreement between raters. Exact agreement is defined as scores that are exactly the same. Adjacent agreement is defined as scores differing by 1 point. In addition, Cohen's kappa is provided as a measure of agreement between the raters. Cohen's kappa (Cohen, 1960) is commonly used to summarize the agreement between raters. It is computed as (Brennan & Prediger, 1981)

$$\kappa = \frac{\sum P_{ii} - \sum P_{i \cdot} \cdot P_{\cdot i}}{1 - \sum P_{i \cdot} \cdot P_{\cdot i}}$$

where $\sum P_{ii}$ is the observed proportion of agreement and $\sum P_{i \cdot} \cdot P_{\cdot i}$ is the chance proportion of agreement. Tables D-1 through D-6 show the rater reliability indices for all 2006 constructed response items by content area.

Rater Severity/Leniency Study

In addition to examining rater reliability measures within a given administration year, CTB also conducts a rater severity study across years. Rater severity or leniency is defined as the extent to which scores assigned by raters across years are systematically offset. The study entails sampling student responses from previous administrations, having a representative group of raters from the current administration score them, and comparing the scores against the scores assigned by the previous raters. The following specifications describe the rater severity study in detail.

1. In 2006, a rater severity study was done using constructed response items that are repeated from 2004.
2. Random samples of student responses were selected from the 2004 CSAP tests for tests where repeating items are present:
 - A random sample of approximately 1000 students was selected for English Reading and English Writing assessments.
 - A random sample of around 250 students was selected for Spanish Reading and Spanish Writing assessments.
 - A random sample of approximately 1000 students was selected for Mathematics, grades 5-10 and Science 8.
3. The samples of papers were administered blindly to the 2006 raters during 2006 operational scoring. That is, the raters scoring the papers from a previous administration neither knew that the papers had been scored before, nor that they came from the 2004 data. The items to be rescored were shown to the 2006 raters under their 2006 item numbers. For example, item # 27 in grade 3 Reading in 2005 was previously administered as item # 18 in 2004. Therefore, the 2004 responses to item # 18 were rescored in 2006 as item # 27.
4. The scores from the rescore were then compared with the original scores given to the papers by the raters in 2004.

Table D-7 shows results of the rater severity study, including mean scores from the 2004 administration, mean scores from the 2006 administration, percent of the scores with exact, adjacent, and discrepant agreement, weighted kappa, correlation, and intraclass correlation.

Weighted Kappa, which may be interpreted as the chance-corrected weighted proportional agreement, is relatively smaller for some constructed-response items probably because of relatively smaller exact agreement rate (specifically for the items with 0 or 1 rating) and smaller sample size.

References

Brennan, R.L., & Prediger, D. J., (1981). Coefficient kappa: Some uses, misuses and alternatives. *Educational and Psychological Measurement*, 41, 687-699.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Table D-1. Rater Reliability Statistics -- Mathematics

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
3	4	2406	2	1.12	1.11	0.93	0.93	0.96	93.4	6.3	99.7
3	9	2416	2	1.49	1.49	0.68	0.69	0.92	93.2	6.4	99.6
3	17	2410	2	1.39	1.39	0.80	0.80	0.93	92.4	7.2	99.6
3	20	2419	3	1.65	1.65	1.02	1.02	0.95	91.2	8.5	99.7
3	24	2410	2	0.94	0.93	0.74	0.75	0.93	92.2	7.8	100.0
3	27	2413	2	1.44	1.44	0.78	0.78	0.97	97.2	2.6	99.8
3	34	2419	2	1.56	1.56	0.65	0.64	0.94	95.6	4.3	99.9
3	40	2417	3	1.71	1.70	1.06	1.06	0.96	91.4	8.2	99.6
4	5	1931	2	0.96	0.96	0.62	0.61	0.89	92.4	7.4	99.9
4	11	1931	2	1.52	1.52	0.73	0.74	0.91	90.7	9.2	99.9
4	15	1927	3	2.25	2.26	0.78	0.79	0.83	81.1	18.3	99.3
4	18	1937	3	1.71	1.71	0.92	0.92	0.92	88.9	10.5	99.3
4	23	1937	4	2.67	2.65	1.22	1.23	0.93	85.3	13.1	98.4
4	27	1933	2	1.09	1.08	0.71	0.72	0.89	89.3	10.7	100.0
4	32	1903	2	1.50	1.49	0.76	0.77	0.86	88.3	9.9	98.3
4	36	1932	3	1.14	1.15	1.14	1.15	0.93	87.7	10.0	97.8
4	41	1917	3	0.95	0.94	0.92	0.92	0.86	82.7	14.9	97.7
4	46	1929	4	2.83	2.84	1.17	1.18	0.90	76.1	22.6	98.7
4	49	1929	2	1.55	1.55	0.59	0.59	0.85	89.5	10.5	100.0
4	54	1931	3	1.97	1.95	0.99	0.99	0.96	93.2	6.6	99.8
4	59	1934	2	1.65	1.65	0.60	0.60	0.93	95.5	4.3	99.8
4	63	1929	3	1.83	1.84	1.18	1.18	0.89	79.3	17.4	96.6
4	69	1929	4	1.69	1.68	1.31	1.33	0.89	74.0	22.3	96.3
5	6	1948	2	1.34	1.34	0.67	0.67	0.92	92.9	7.0	99.9
5	13	1942	3	2.17	2.16	0.81	0.82	0.87	83.5	16.2	99.7
5	15	1944	2	0.93	0.93	0.96	0.96	0.95	92.7	6.8	99.5
5	19	1942	3	1.84	1.84	1.00	1.00	0.84	75.9	22.0	97.8
5	23	1948	4	3.31	3.31	0.99	0.99	0.89	82.6	16.1	98.7
5	27	1944	2	1.35	1.35	0.87	0.87	0.91	88.3	11.0	99.3
5	34	1938	3	2.13	2.13	1.07	1.08	0.94	90.0	9.3	99.3
5	35	1874	2	1.77	1.76	0.55	0.56	0.80	92.1	6.6	98.6
5	40	1942	3	1.68	1.67	1.01	1.02	0.89	80.6	18.3	99.0
5	46	1946	4	2.97	2.99	1.11	1.10	0.86	74.5	22.6	97.0
5	49	1951	2	1.45	1.46	0.84	0.84	0.97	97.2	2.4	99.5
5	53	1940	3	2.00	1.99	1.08	1.07	0.86	78.8	17.8	96.6
5	57	1942	2	1.41	1.42	0.74	0.75	0.84	83.6	16.1	99.7
5	66	1946	3	2.07	2.07	1.10	1.10	0.89	80.9	16.3	97.3
5	69	1947	4	2.61	2.61	1.25	1.25	0.94	83.7	15.5	99.2

Table D-1. Rater Reliability Statistics – Mathematics (continued)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
6	4	1530	2	1.67	1.67	0.56	0.56	0.90	94.6	5.0	99.6
6	7	1530	3	2.44	2.45	0.77	0.76	0.85	86.9	11.8	98.7
6	11	1531	2	0.55	0.57	0.77	0.78	0.91	91.4	7.9	99.3
6	16	1509	3	1.39	1.40	1.23	1.22	0.96	89.5	9.7	99.2
6	20	1533	4	1.98	1.97	1.00	1.01	0.89	79.7	19.6	99.3
6	26	1536	2	1.76	1.76	0.44	0.45	0.70	88.9	11.0	99.9
6	29	1532	3	2.08	2.08	1.06	1.06	0.95	89.4	10.3	99.7
6	33	1535	2	1.29	1.28	0.83	0.84	0.88	86.1	13.0	99.1
6	37	1535	3	0.93	0.93	0.99	0.99	0.89	89.5	7.2	96.7
6	40	1534	4	3.28	3.29	1.06	1.04	0.87	78.8	19.0	97.8
6	43	1536	2	1.47	1.48	0.72	0.72	0.96	96.2	3.7	99.9
6	47	1519	2	0.74	0.73	0.77	0.76	0.83	85.7	12.4	98.0
6	48	1535	3	1.97	1.99	0.97	0.97	0.92	85.8	13.7	99.5
6	57	1534	3	1.92	1.92	0.98	0.97	0.97	96.7	2.3	99.0
6	60	1534	4	1.72	1.72	1.34	1.32	0.92	79.6	18.1	97.7
7	3	1447	2	0.86	0.84	0.86	0.86	0.94	92.1	7.5	99.6
7	7	1398	3	1.45	1.46	1.21	1.21	0.89	76.7	20.7	97.4
7	13	1389	2	0.94	0.96	0.90	0.90	0.82	78.8	18.5	97.3
7	15	1447	3	1.61	1.59	1.04	1.05	0.91	81.6	17.8	99.4
7	20	1427	4	1.34	1.35	1.48	1.48	0.94	82.0	16.1	98.0
7	25	1454	2	1.42	1.41	0.77	0.78	0.92	91.6	8.2	99.8
7	29	1443	3	0.96	0.96	0.81	0.82	0.82	77.3	22.1	99.5
7	33	1452	2	1.73	1.73	0.45	0.45	0.79	91.5	8.5	100.0
7	35	1455	3	1.89	1.89	0.91	0.93	0.90	85.2	13.8	99.0
7	40	1420	4	0.68	0.68	1.18	1.18	0.92	83.1	15.1	98.2
7	44	1444	2	1.19	1.20	0.79	0.80	0.89	87.5	12.2	99.7
7	48	1443	3	0.75	0.75	0.87	0.87	0.88	84.6	14.4	99.0
7	52	1441	2	1.21	1.19	0.79	0.78	0.73	79.3	16.7	96.0
7	55	1423	3	0.47	0.49	0.75	0.79	0.82	85.8	12.0	97.8
7	60	1440	4	2.15	2.16	1.41	1.40	0.88	69.6	25.7	95.3
8	4	1470	3	2.11	2.09	0.85	0.87	0.83	78.9	19.8	98.7
8	6	1405	2	1.12	1.09	0.89	0.90	0.82	79.6	17.4	97.0
8	8	1461	2	1.21	1.23	0.79	0.78	0.82	85.5	12.1	97.5
8	13	1455	3	0.64	0.63	0.92	0.92	0.95	94.6	4.4	99.0
8	20	1450	4	1.40	1.39	1.50	1.49	0.94	80.8	16.3	97.1
8	25	1451	2	0.16	0.15	0.50	0.49	0.95	98.1	1.7	99.8
8	28	1446	3	0.79	0.79	0.81	0.82	0.89	86.7	13.1	99.8
8	31	1457	2	0.87	0.88	0.71	0.71	0.90	90.6	9.3	99.9
8	35	1454	3	1.59	1.60	1.15	1.15	0.95	88.4	11.1	99.5
8	40	1464	4	2.62	2.62	0.89	0.90	0.85	80.8	17.8	98.6
8	44	1414	2	0.80	0.79	0.88	0.89	0.91	86.5	13.1	99.6
8	46	1441	3	1.28	1.28	1.16	1.15	0.91	81.5	16.6	98.1
8	49	1454	2	0.86	0.86	0.84	0.84	0.82	76.0	23.7	99.7
8	55	1448	3	1.10	1.10	0.95	0.96	0.76	64.9	32.0	96.9
8	60	1445	4	1.83	1.81	1.64	1.62	0.96	83.0	14.7	97.7

Table D-1. Rater Reliability Statistics – Mathematics (concluded)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
9	4	1377	2	0.49	0.48	0.74	0.74	0.96	95.5	4.4	99.9
9	8	1349	3	1.70	1.72	1.03	1.02	0.93	87.4	12.0	99.4
9	12	1400	2	0.99	1.00	0.83	0.83	0.82	77.9	21.4	99.3
9	15	1368	3	0.49	0.49	0.83	0.82	0.91	93.2	5.2	98.4
9	20	1382	4	1.97	1.96	1.08	1.10	0.88	77.1	21.4	98.5
9	24	1400	2	0.94	0.94	0.72	0.72	0.95	95.4	4.6	99.9
9	27	1371	3	0.76	0.78	0.84	0.85	0.87	82.6	17.0	99.6
9	34	1376	3	1.47	1.47	1.00	0.99	0.82	69.2	29.1	98.3
9	36	1388	2	0.77	0.77	0.76	0.76	0.96	95.2	4.8	99.9
9	40	1384	4	0.66	0.67	1.06	1.06	0.93	85.8	14.0	99.7
9	43	1342	2	0.49	0.48	0.73	0.73	0.90	89.3	10.7	99.9
9	46	1385	3	0.84	0.84	1.06	1.06	0.92	87.4	11.0	98.3
9	50	1385	2	0.96	0.96	0.73	0.74	0.95	94.9	4.9	99.8
9	54	1382	3	1.24	1.25	0.89	0.88	0.72	68.1	27.9	96.0
9	60	1365	4	1.41	1.41	0.94	0.97	0.75	67.3	29.1	96.3
10	4	1348	2	1.46	1.45	0.81	0.81	0.87	89.2	8.8	97.9
10	9	1323	3	0.61	0.62	0.90	0.92	0.92	88.4	11.2	99.6
10	15	1352	2	0.85	0.83	0.70	0.70	0.84	84.5	15.5	100.0
10	18	1336	3	0.70	0.71	0.90	0.91	0.90	91.0	6.9	97.9
10	20	1307	4	0.33	0.35	0.67	0.73	0.58	76.3	19.4	95.6
10	25	1341	2	0.76	0.77	0.80	0.80	0.92	89.6	10.2	99.9
10	28	1348	3	1.29	1.31	1.03	1.03	0.93	91.5	6.5	97.9
10	32	1340	2	0.84	0.86	0.60	0.60	0.76	85.0	14.3	99.3
10	34	1307	3	1.86	1.84	1.00	1.03	0.90	81.6	17.5	99.1
10	40	1343	4	1.73	1.72	1.48	1.49	0.91	80.9	13.5	94.4
10	45	1348	2	0.64	0.66	0.83	0.83	0.94	92.7	7.1	99.9
10	48	1342	3	0.79	0.80	1.11	1.11	0.94	87.0	12.7	99.7
10	54	1348	2	1.26	1.27	0.51	0.52	0.74	86.8	13.1	99.9
10	57	1343	3	0.91	0.92	1.06	1.05	0.93	87.0	11.8	98.8
10	60	1347	4	1.52	1.52	0.94	0.92	0.79	71.4	26.1	97.6

Table D-2. Rater Reliability Statistics -- Reading

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
3	1	2577	3	1.15	1.16	1.24	1.25	0.78	77.0	15.0	91.9
3	7	2575	2	1.29	1.29	0.70	0.70	0.85	86.4	13.2	99.5
3	8	2539	2	1.17	1.17	0.85	0.85	0.95	92.9	6.9	99.8
3	14	2513	2	0.81	0.80	0.88	0.88	0.96	94.7	5.1	99.8
3	22	2574	3	1.48	1.49	1.36	1.35	0.87	90.1	3.9	93.9
3	24	2567	3	1.54	1.53	1.33	1.33	0.99	97.0	2.6	99.6
3	27	2567	2	0.95	0.95	0.84	0.83	0.94	92.7	7.1	99.7
3	33	2557	3	1.89	1.90	1.14	1.14	0.91	81.9	16.6	98.6
4	4	1731	3	1.38	1.40	1.27	1.26	0.92	91.2	4.2	95.4
4	14	1731	2	0.55	0.58	0.85	0.86	0.87	90.8	6.1	96.9
4	17	1698	4	1.72	1.69	1.23	1.24	0.81	63.6	30.2	93.8
4	21	1708	2	0.64	0.66	0.69	0.69	0.86	88.2	11.2	99.5
4	24	1708	2	0.92	0.92	0.94	0.94	0.97	96.0	3.4	99.4
4	27	1718	3	1.48	1.47	1.32	1.32	0.71	71.8	16.3	88.1
4	33	1729	2	1.50	1.49	0.67	0.67	0.81	85.3	13.9	99.3
4	38	1719	3	1.49	1.49	1.03	1.02	0.85	84.3	12.6	96.9
4	44	1718	2	1.18	1.18	0.87	0.86	0.78	79.7	16.4	96.1
4	97	1725	2	0.72	0.72	0.74	0.73	0.84	83.7	15.9	99.5
4	103	1720	3	2.41	2.40	0.93	0.93	0.87	82.9	15.9	98.7
4	109	1704	2	0.69	0.71	0.74	0.76	0.78	80.1	18.4	98.5
4	115	1725	3	1.35	1.35	0.95	0.95	0.94	89.2	10.6	99.8
4	116	1700	2	0.82	0.82	0.89	0.88	0.93	90.5	9.0	99.5
5	4	1698	2	1.67	1.68	0.60	0.60	0.86	91.0	8.5	99.5
5	11	1681	2	0.78	0.77	0.85	0.86	0.87	84.3	14.4	98.7
5	12	1686	2	0.72	0.71	0.73	0.73	0.81	81.5	17.8	99.3
5	16	1690	2	0.86	0.87	0.61	0.62	0.86	90.2	9.5	99.7
5	19	1687	3	1.96	1.96	0.97	0.96	0.93	87.1	12.6	99.8
5	25	1658	4	1.93	1.93	1.22	1.23	0.84	68.3	27.4	95.8
5	31	1689	2	0.54	0.52	0.71	0.69	0.67	73.8	24.1	97.9
5	39	1676	3	0.59	0.58	0.94	0.95	0.42	62.4	22.8	85.2
5	44	1691	3	1.40	1.40	1.06	1.06	0.89	87.8	9.5	97.3
5	49	1680	3	1.10	1.11	0.97	0.98	0.80	76.5	18.6	95.1
5	50	1688	2	1.26	1.28	0.82	0.82	0.77	78.6	18.4	96.9
5	93	1685	2	1.38	1.38	0.88	0.88	0.93	92.5	6.5	99.0
5	94	1687	2	1.30	1.29	0.77	0.77	0.93	92.8	7.1	99.8
5	106	1683	2	1.33	1.33	0.81	0.81	0.93	91.3	8.6	99.9
5	107	1679	2	1.25	1.25	0.72	0.71	0.87	87.9	11.6	99.5

Table D-2. Rater Reliability Statistics – Reading (continued)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
6	4	1671	2	1.74	1.74	0.55	0.55	0.82	89.4	10.5	99.8
6	11	1652	2	0.94	0.97	0.86	0.88	0.83	80.3	17.6	97.8
6	12	1656	4	2.99	2.98	1.32	1.36	0.83	72.0	19.6	91.7
6	13	1663	2	1.12	1.11	0.88	0.87	0.79	78.2	18.3	96.5
6	21	1631	2	1.11	1.12	0.93	0.93	0.88	87.7	9.7	97.4
6	22	1609	2	1.33	1.35	0.83	0.83	0.90	87.1	12.5	99.6
6	35	1666	2	0.60	0.60	0.70	0.70	0.69	73.6	25.0	98.6
6	43	1631	3	1.00	0.97	1.09	1.08	0.50	51.6	32.0	83.6
6	44	1664	2	1.42	1.42	0.72	0.72	0.96	96.1	3.9	100.0
6	45	1659	2	1.68	1.69	0.67	0.66	0.97	97.8	2.0	99.8
6	98	1653	3	1.00	0.97	1.06	1.05	0.84	75.7	21.2	96.9
6	99	1646	2	1.07	1.07	0.86	0.86	0.93	90.7	9.1	99.8
6	111	1638	2	1.03	1.03	0.84	0.84	0.87	84.0	15.3	99.3
6	114	1642	3	1.06	1.08	0.96	0.96	0.85	77.7	20.8	98.4
7	4	1391	2	1.28	1.26	0.79	0.81	0.81	78.7	20.1	98.8
7	9	1398	2	1.36	1.33	0.73	0.72	0.70	73.3	25.1	98.4
7	10	1396	2	1.01	1.01	0.82	0.81	0.65	71.3	22.8	94.1
7	11	1391	2	1.43	1.42	0.80	0.81	0.94	94.2	5.0	99.2
7	19	1386	2	1.05	1.05	0.95	0.94	0.90	90.0	7.4	97.5
7	20	1378	2	1.39	1.40	0.83	0.83	0.93	91.7	8.0	99.7
7	36	1387	3	0.83	0.83	0.85	0.83	0.86	81.5	18.0	99.4
7	37	1400	2	1.04	1.04	0.71	0.69	0.80	82.0	17.6	99.6
7	89	1362	3	0.95	0.89	0.88	0.83	0.71	67.8	28.9	96.7
7	99	1384	3	1.42	1.42	1.06	1.04	0.83	69.6	27.8	97.3
7	100	1379	2	0.94	0.93	0.80	0.81	0.83	80.8	18.4	99.2
7	103	1377	3	1.09	1.08	0.93	0.94	0.82	74.5	23.3	97.8
7	109	1385	2	1.66	1.68	0.58	0.56	0.70	82.2	17.1	99.4
7	115	1374	2	1.09	1.10	0.78	0.79	0.82	81.8	17.0	98.8
8	9	1389	2	0.97	0.99	0.78	0.79	0.73	73.2	24.4	97.6
8	13	1414	3	1.60	1.57	1.10	1.08	0.73	59.4	34.5	93.9
8	14	1420	2	1.72	1.73	0.56	0.55	0.82	89.7	9.9	99.7
8	22	1394	2	1.39	1.41	0.85	0.84	0.81	83.1	13.4	96.5
8	29	1381	2	0.72	0.70	0.85	0.84	0.69	73.6	20.4	94.0
8	30	1366	2	1.11	1.09	0.83	0.83	0.74	73.2	24.0	97.2
8	37	1391	3	0.88	0.87	0.87	0.86	0.87	82.3	17.0	99.4
8	38	1414	2	1.04	1.04	0.71	0.71	0.81	82.7	16.8	99.5
8	87	1390	3	0.92	0.93	0.83	0.84	0.68	63.4	33.9	97.3
8	99	1369	2	0.70	0.70	0.62	0.62	0.84	88.5	11.3	99.7
8	100	1355	2	0.38	0.38	0.61	0.61	0.72	84.4	14.0	98.3
8	106	1389	2	1.26	1.27	0.84	0.83	0.91	88.4	11.2	99.6
8	107	1306	3	1.36	1.34	1.19	1.20	0.87	72.9	24.6	97.5
8	114	1295	3	1.78	1.78	1.31	1.31	0.95	87.9	11.0	98.9

Table D-2. Rater Reliability Statistics – Reading (concluded)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
9	4	1553	3	1.71	1.68	1.10	1.10	0.73	60.1	33.0	93.1
9	5	1560	2	1.81	1.81	0.48	0.48	0.80	91.9	7.8	99.7
9	12	1547	3	1.94	1.98	0.80	0.79	0.71	70.3	27.7	97.9
9	26	1503	3	1.85	1.83	1.19	1.19	0.81	67.6	26.7	94.3
9	27	1524	2	1.08	1.05	0.76	0.75	0.82	81.2	18.1	99.3
9	38	1502	2	0.97	0.98	0.77	0.77	0.58	60.5	36.2	96.7
9	43	1497	3	1.73	1.70	1.09	1.09	0.80	65.4	30.8	96.2
9	44	1498	3	2.12	2.13	0.91	0.91	0.71	64.4	31.8	96.2
9	93	1517	2	0.80	0.80	0.62	0.64	0.82	87.0	12.6	99.5
9	94	1495	2	0.49	0.50	0.69	0.70	0.74	80.7	17.4	98.1
9	100	1497	3	1.53	1.51	1.22	1.23	0.73	65.3	25.0	90.2
9	104	1519	2	1.00	0.99	0.74	0.75	0.61	66.2	30.9	97.0
9	107	1494	3	2.11	2.10	1.03	1.02	0.96	93.8	5.6	99.5
9	118	1526	2	0.97	0.98	0.69	0.68	0.52	60.0	38.3	98.3
10	8	1358	2	0.69	0.68	0.73	0.72	0.78	80.3	18.6	99.0
10	9	1337	3	2.23	2.21	1.12	1.12	0.87	83.9	11.4	95.4
10	20	1327	3	1.57	1.57	1.24	1.24	0.96	92.6	6.0	98.6
10	28	1308	3	1.79	1.81	0.89	0.89	0.71	61.2	36.6	97.8
10	31	1326	3	0.96	0.95	0.93	0.92	0.83	78.4	19.7	98.1
10	35	1341	2	0.92	0.93	0.75	0.75	0.65	72.0	24.3	96.3
10	41	1228	4	1.46	1.46	1.25	1.25	0.92	84.8	12.2	97.0
10	43	1356	3	1.96	1.95	0.80	0.78	0.72	68.8	30.0	98.8
10	49	1361	3	2.34	2.33	0.91	0.92	0.85	81.2	16.9	98.1
10	50	1336	3	1.39	1.41	0.89	0.89	0.69	68.7	25.2	93.9
10	96	1309	3	1.44	1.46	1.15	1.15	0.65	61.9	27.1	89.0
10	102	1343	2	0.96	0.97	0.74	0.73	0.63	67.0	30.5	97.5
10	106	1338	3	1.05	1.05	1.02	1.03	0.89	82.0	16.7	98.7
10	109	1302	2	0.79	0.79	0.85	0.87	0.78	83.2	11.8	95.0

Table D-3. Rater Reliability Statistics -- Writing

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
3	5	2013	1	0.93	0.92	0.26	0.27	0.83	97.5	2.5	100.0
3	5	2013	1	0.72	0.72	0.45	0.45	0.94	97.5	2.5	100.0
3	5	2013	1	0.89	0.89	0.31	0.31	0.93	98.6	1.4	100.0
3	5	2012	1	0.06	0.06	0.24	0.23	0.88	98.6	1.4	100.0
3	5	2013	1	0.82	0.82	0.38	0.39	0.95	98.6	1.4	100.0
3	5	2013	1	0.70	0.70	0.46	0.46	0.96	98.5	1.5	100.0
3	12	2011	2	1.02	1.01	0.23	0.22	0.22	92.3	7.6	99.9
3	12	2011	1	0.98	0.98	0.14	0.14	0.50	98.1	1.9	100.0
3	12	2011	1	0.80	0.80	0.40	0.40	0.57	86.2	13.8	100.0
3	12	2011	1	0.95	0.96	0.22	0.20	0.38	94.7	5.3	100.0
3	30	2012	2	1.15	1.17	0.43	0.44	0.45	79.3	20.7	100.0
3	30	2012	1	0.97	0.97	0.17	0.17	0.56	97.5	2.5	100.0
3	30	2012	1	0.85	0.85	0.36	0.36	0.53	88.1	11.9	100.0
3	30	2012	1	0.92	0.91	0.28	0.28	0.42	90.8	9.2	100.0
3	39	2009	2	1.07	1.07	0.42	0.43	0.46	81.9	17.7	99.7
3	39	2009	1	0.94	0.94	0.23	0.24	0.46	94.0	6.0	100.0
3	39	2009	1	0.80	0.80	0.40	0.40	0.54	85.3	14.7	100.0
3	39	2009	1	0.89	0.90	0.31	0.30	0.41	89.0	11.1	100.0
4	1	1745	1	0.96	0.96	0.20	0.19	0.83	98.7	1.3	100.0
4	2	1728	1	0.55	0.56	0.50	0.50	0.94	97.2	2.8	100.0
4	2	1729	1	0.31	0.31	0.46	0.46	0.96	98.5	1.5	100.0
4	2	1729	1	0.41	0.41	0.49	0.49	0.96	98.2	1.9	100.0
4	2	1730	1	0.71	0.71	0.45	0.45	0.92	96.7	3.4	100.0
4	2	1729	1	0.89	0.89	0.32	0.32	0.84	96.8	3.2	100.0
4	2	1729	1	0.27	0.27	0.44	0.44	0.95	98.0	2.0	100.0
4	3	1734	4	2.85	2.82	0.82	0.81	0.71	67.1	31.6	98.7
4	3	1735	4	2.80	2.77	0.78	0.76	0.61	60.3	37.7	98.0
4	3	1735	2	1.87	1.86	0.37	0.37	0.43	85.7	14.0	99.7
4	53	1695	4	2.17	2.20	0.88	0.89	0.68	63.0	33.5	96.4
4	76	1725	4	2.37	2.38	0.74	0.75	0.61	63.5	34.6	98.1
4	95	1719	4	2.13	2.16	0.86	0.89	0.65	59.2	36.7	95.9
5	1	1700	1	0.98	0.98	0.12	0.13	0.59	98.7	1.3	100.0
5	2	1688	1	0.06	0.06	0.24	0.23	0.82	98.0	2.0	100.0
5	2	1688	1	0.64	0.64	0.48	0.48	0.90	95.4	4.6	100.0
5	2	1688	1	0.44	0.44	0.50	0.50	0.97	98.5	1.5	100.0
5	2	1688	1	0.65	0.65	0.48	0.48	0.95	97.5	2.5	100.0
5	2	1688	1	0.50	0.50	0.50	0.50	0.97	98.6	1.4	100.0
5	2	1688	1	0.48	0.48	0.50	0.50	0.93	96.4	3.6	100.0
5	3	1704	4	2.21	2.24	1.18	1.17	0.68	59.4	30.5	89.9
5	3	1704	4	2.44	2.45	0.87	0.87	0.68	60.6	36.6	97.1
5	3	1704	2	1.79	1.80	0.43	0.44	0.52	82.6	17.3	99.8
5	62	1689	4	2.47	2.50	0.74	0.74	0.59	61.9	36.2	98.1
5	92	1685	4	2.35	2.36	0.82	0.82	0.67	62.4	35.6	98.0
5	117	1687	4	2.42	2.41	0.71	0.70	0.62	66.0	33.0	99.0

Table D-3. Rater Reliability Statistics – Writing (continued)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
6	1	1667	1	0.99	1.00	0.07	0.04	0.33	99.5	0.5	100.0
6	2	1661	1	0.68	0.69	0.47	0.46	0.96	98.3	1.7	100.0
6	2	1660	1	0.59	0.59	0.49	0.49	0.96	98.2	1.8	100.0
6	2	1660	1	0.72	0.72	0.45	0.45	0.95	98.1	1.9	100.0
6	2	1661	1	0.66	0.66	0.47	0.48	0.96	98.3	1.8	100.0
6	2	1662	1	0.24	0.24	0.43	0.43	0.96	98.5	1.5	100.0
6	2	1662	1	0.83	0.83	0.37	0.37	0.95	98.7	1.3	100.0
6	3	1671	4	2.42	2.42	0.85	0.80	0.67	63.0	35.3	98.3
6	3	1671	4	2.41	2.41	0.81	0.77	0.68	63.7	34.9	98.6
6	3	1671	2	1.84	1.84	0.37	0.37	0.50	86.3	13.7	100.0
6	32	1536	4	2.48	2.44	0.78	0.78	0.66	63.3	35.3	98.6
6	78	1647	4	2.54	2.53	0.74	0.74	0.65	65.5	33.3	98.8
6	97	1638	4	2.33	2.31	0.72	0.72	0.61	64.5	34.1	98.7
7	1	1411	1	1.00	1.00	0.07	0.07	0.86	99.9	0.1	100.0
7	2	1394	1	0.38	0.38	0.49	0.49	0.91	95.9	4.1	100.0
7	2	1394	1	0.47	0.47	0.50	0.50	0.93	96.3	3.7	100.0
7	2	1394	1	0.37	0.37	0.48	0.48	0.95	97.7	2.3	100.0
7	2	1394	1	0.45	0.45	0.50	0.50	0.98	98.9	1.1	100.0
7	2	1394	1	0.81	0.81	0.39	0.39	0.96	98.8	1.2	100.0
7	2	1394	1	0.44	0.45	0.50	0.50	0.98	98.9	1.2	100.0
7	3	1403	4	2.76	2.77	0.81	0.84	0.66	62.8	35.1	97.9
7	3	1403	4	2.75	2.75	0.77	0.80	0.66	61.7	37.0	98.7
7	3	1403	2	1.88	1.87	0.33	0.34	0.39	86.5	13.4	99.9
7	65	1386	4	2.62	2.61	0.80	0.83	0.67	61.7	36.6	98.3
7	88	1379	4	2.51	2.50	0.72	0.70	0.66	67.5	31.9	99.4
7	118	1368	4	2.43	2.43	0.75	0.77	0.68	67.0	31.9	99.0
8	1	1439	1	0.99	0.99	0.11	0.10	0.82	99.6	0.4	100.0
8	2	1418	1	0.63	0.63	0.48	0.48	0.96	98.1	1.9	100.0
8	2	1418	1	0.62	0.62	0.48	0.49	0.91	95.8	4.2	100.0
8	2	1418	1	0.25	0.25	0.43	0.43	0.94	97.7	2.3	100.0
8	2	1418	1	0.53	0.53	0.50	0.50	0.80	90.1	9.9	100.0
8	2	1416	1	0.26	0.26	0.44	0.44	0.98	99.1	0.9	100.0
8	2	1418	1	0.58	0.58	0.49	0.49	0.95	97.7	2.3	100.0
8	3	1427	4	2.53	2.52	0.73	0.72	0.70	69.6	30.1	99.7
8	3	1428	4	2.52	2.53	0.72	0.70	0.68	68.7	31.0	99.7
8	3	1428	2	1.89	1.88	0.32	0.32	0.40	87.9	12.0	99.9
8	21	1408	4	2.54	2.56	0.76	0.75	0.66	64.9	33.7	98.6
8	69	1393	4	2.62	2.59	0.78	0.80	0.65	63.3	34.6	97.9
8	86	1399	4	2.49	2.46	0.67	0.70	0.58	63.2	36.1	99.3

Table D-3. Rater Reliability Statistics – Writing (concluded)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
9	1	1579	1	0.99	0.99	0.11	0.11	0.92	99.8	0.2	100.0
9	2	1548	1	0.81	0.81	0.39	0.39	0.98	99.3	0.7	100.0
9	2	1548	1	0.91	0.91	0.28	0.28	0.94	99.0	1.0	100.0
9	2	1548	1	0.42	0.42	0.49	0.49	0.94	97.2	2.8	100.0
9	2	1548	1	0.81	0.81	0.39	0.39	0.90	97.1	2.9	100.0
9	2	1548	1	0.47	0.48	0.50	0.50	0.95	97.4	2.7	100.0
9	2	1548	1	0.87	0.87	0.34	0.33	0.83	96.3	3.8	100.0
9	3	1564	4	2.30	2.29	0.67	0.67	0.71	75.5	24.2	99.7
9	3	1564	4	2.31	2.32	0.66	0.65	0.64	70.1	29.5	99.6
9	3	1564	2	1.81	1.81	0.40	0.39	0.49	84.3	15.6	99.9
9	24	1544	4	2.53	2.53	0.75	0.78	0.73	68.9	30.8	99.7
9	70	1552	4	2.41	2.41	0.69	0.69	0.68	70.4	29.5	99.8
9	92	1536	4	2.41	2.39	0.72	0.72	0.72	71.6	28.3	99.9
10	1	1406	1	0.98	0.98	0.14	0.14	0.96	99.9	0.1	100.0
10	2	1374	1	0.59	0.59	0.49	0.49	0.95	97.5	2.5	100.0
10	2	1374	1	0.66	0.65	0.48	0.48	0.95	97.7	2.3	100.0
10	2	1374	1	0.37	0.37	0.48	0.48	0.93	96.5	3.5	100.0
10	2	1374	1	0.49	0.50	0.50	0.50	0.97	98.3	1.8	100.0
10	2	1374	1	0.58	0.58	0.49	0.49	0.97	98.3	1.7	100.0
10	2	1374	1	0.83	0.83	0.38	0.38	0.96	99.0	1.0	100.0
10	3	1386	4	2.48	2.48	0.62	0.64	0.65	73.8	25.7	99.5
10	3	1386	4	2.47	2.48	0.64	0.65	0.62	70.4	28.7	99.1
10	3	1386	2	1.84	1.85	0.36	0.36	0.45	85.6	14.4	100.0
10	73	1363	4	2.63	2.62	0.68	0.68	0.70	73.1	26.8	99.9
10	95	1347	4	2.59	2.61	0.71	0.70	0.70	70.5	29.3	99.8
10	115	1357	4	2.50	2.50	0.71	0.73	0.71	71.0	28.5	99.6

Table D-4. Rater Reliability Statistics -- Science

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
5	5	2238	2	1.37	1.37	0.80	0.79	0.90	89.4	10.0	99.4
5	21	2248	1	0.72	0.73	0.45	0.45	0.72	89.0	11.0	100.0
5	22	2235	1	0.42	0.42	0.49	0.49	0.91	95.8	4.2	100.0
5	23	2232	2	1.00	1.01	0.76	0.76	0.89	88.6	11.0	99.6
5	24	2224	2	0.24	0.24	0.49	0.49	0.70	87.4	12.0	99.4
5	25	2229	1	0.56	0.55	0.50	0.50	0.87	93.8	6.2	100.0
5	29	2240	2	0.74	0.72	0.87	0.86	0.76	78.3	17.2	95.5
5	34	2231	2	0.51	0.51	0.74	0.74	0.89	89.6	9.8	99.3
5	38	2241	2	0.97	0.97	0.27	0.27	0.87	98.2	1.7	100.0
5	41	2235	1	0.41	0.41	0.49	0.49	0.66	83.7	16.3	100.0
5	43	2246	2	1.52	1.52	0.59	0.59	0.95	96.7	3.3	100.0
5	46	2226	2	1.30	1.29	0.83	0.83	0.81	79.3	18.7	98.0
5	53	2230	1	0.25	0.25	0.43	0.43	0.90	96.1	3.9	100.0
5	54	2248	2	1.82	1.82	0.40	0.41	0.91	97.8	2.0	99.7
5	57	2240	2	1.55	1.54	0.62	0.62	0.86	90.0	9.9	99.8
5	64	2243	2	1.71	1.72	0.53	0.53	0.83	90.6	9.4	100.0
5	71	2245	2	0.84	0.83	0.69	0.69	0.80	83.2	16.0	99.1
5	75	2227	2	0.69	0.70	0.84	0.85	0.88	84.1	15.4	99.5
8	9	2331	2	0.92	0.93	0.82	0.82	0.92	88.9	11.1	100.0
8	13	2430	1	0.06	0.07	0.24	0.25	0.94	99.3	0.7	100.0
8	15	2399	1	0.47	0.46	0.50	0.50	0.65	82.5	17.6	100.0
8	20	2411	1	0.53	0.53	0.50	0.50	0.83	91.5	8.5	100.0
8	21	2419	1	0.67	0.67	0.47	0.47	0.69	86.2	13.9	100.0
8	22	2370	2	1.23	1.22	0.82	0.83	0.90	87.9	11.5	99.4
8	23	2390	1	0.60	0.61	0.49	0.49	0.76	88.3	11.7	100.0
8	24	2418	4	2.24	2.25	1.49	1.49	0.85	66.8	24.5	91.4
8	25	2381	2	1.39	1.38	0.75	0.75	0.89	88.4	11.4	99.8
8	34	2370	2	0.52	0.51	0.55	0.55	0.86	91.4	8.7	100.0
8	36	2426	2	0.86	0.86	0.51	0.52	0.72	85.2	14.7	99.9
8	42	2425	2	1.31	1.31	0.69	0.68	0.88	89.5	10.4	99.9
8	47	2353	2	0.66	0.65	0.60	0.58	0.61	73.6	26.3	99.8
8	50	2154	2	0.38	0.40	0.66	0.68	0.81	85.9	13.1	98.9
8	54	2373	2	0.84	0.84	0.66	0.68	0.82	84.8	14.9	99.7
8	56	2402	3	0.93	0.92	0.81	0.81	0.85	83.0	16.3	99.3
8	60	2379	2	0.49	0.49	0.67	0.67	0.86	88.1	11.7	99.8
8	62	2329	2	0.54	0.53	0.65	0.65	0.97	97.6	2.4	100.0
8	64	2317	1	0.44	0.44	0.50	0.50	0.84	92.4	7.6	100.0
8	65	2409	3	1.28	1.29	1.13	1.13	0.94	86.8	12.7	99.5
8	70	2326	1	0.16	0.16	0.36	0.36	0.72	92.6	7.4	100.0
8	77	2422	1	0.83	0.83	0.37	0.37	0.97	99.3	0.7	100.0
8	81	2382	1	0.64	0.63	0.48	0.48	0.56	79.6	20.4	100.0

Table D-4. Rater Reliability Statistics – Science (concluded)

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
10	12	2235	1	0.64	0.65	0.48	0.48	0.71	86.8	13.2	100.0
10	21	2222	4	2.57	2.57	1.47	1.46	0.84	65.4	26.7	92.1
10	22	2242	1	0.20	0.20	0.40	0.40	0.66	89.1	10.9	100.0
10	23	2187	3	1.86	1.87	1.04	1.05	0.87	78.7	19.7	98.4
10	24	2235	1	0.35	0.35	0.48	0.48	0.76	89.2	10.8	100.0
10	25	2214	1	0.50	0.50	0.50	0.50	0.57	78.5	21.5	100.0
10	26	2193	1	0.50	0.50	0.50	0.50	0.68	84.1	15.9	100.0
10	27	2153	2	0.49	0.48	0.66	0.65	0.54	70.4	26.4	96.8
10	29	2223	2	0.87	0.86	0.78	0.79	0.91	89.6	10.1	99.6
10	34	2044	2	0.70	0.69	0.81	0.81	0.76	77.6	19.5	97.1
10	38	2136	2	0.51	0.50	0.70	0.69	0.92	92.8	7.0	99.8
10	40	2236	2	1.50	1.50	0.64	0.65	0.63	75.0	23.2	98.2
10	46	2233	2	0.33	0.33	0.58	0.58	0.91	94.3	5.5	99.8
10	50	2182	2	0.70	0.70	0.68	0.67	0.89	90.3	9.6	99.9
10	52	2167	1	0.27	0.26	0.44	0.44	0.87	95.1	4.9	100.0
10	58	2108	2	0.31	0.31	0.58	0.57	0.86	91.1	8.8	99.9
10	63	1912	2	0.32	0.31	0.61	0.59	0.77	86.0	13.1	99.1
10	69	2228	1	0.20	0.20	0.40	0.40	0.69	89.9	10.1	100.0
10	71	2107	2	1.37	1.36	0.68	0.68	0.87	88.0	11.8	99.8
10	75	2179	1	0.59	0.59	0.49	0.49	0.71	85.9	14.1	100.0
10	77	2057	2	0.71	0.71	0.73	0.72	0.80	81.0	18.1	99.1
10	78	2120	1	0.19	0.20	0.39	0.40	0.70	90.7	9.3	100.0
10	83	1918	2	0.12	0.12	0.44	0.43	0.97	98.8	1.2	100.0

Table D-5. Rater Reliability Statistics – Spanish Reading

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
3	1	358	3	1.58	1.56	1.26	1.23	0.94	89.4	8.7	98.1
3	2	351	3	1.92	1.83	1.29	1.30	0.89	87.2	8.0	95.2
3	8	360	2	1.19	1.23	0.82	0.82	0.89	88.6	10.0	98.6
3	17	350	2	1.66	1.63	0.69	0.70	0.81	90.0	7.1	97.1
3	25	302	2	1.38	1.41	0.74	0.71	0.78	79.5	19.5	99.0
3	26	299	2	0.67	0.69	0.82	0.84	0.83	81.6	16.7	98.3
3	32	301	3	1.02	1.01	0.84	0.84	0.94	92.7	7.0	99.7
3	36	296	2	1.23	1.23	0.80	0.77	0.82	80.4	18.9	99.3
4	13	78	3	1.15	1.14	1.13	1.09	0.94	85.9	14.1	100.0
4	14	74	2	0.41	0.45	0.55	0.53	0.64	79.7	20.3	100.0
4	28	79	3	2.30	2.28	1.03	1.05	0.96	92.4	7.6	100.0
4	30	79	2	0.37	0.35	0.64	0.64	0.95	96.2	3.8	100.0
4	32	79	3	0.82	0.76	0.93	0.92	0.89	81.0	19.0	100.0
4	33	79	3	1.42	1.34	1.15	1.14	0.81	67.1	29.1	96.2
4	41	74	2	0.49	0.42	0.71	0.64	0.87	87.8	12.2	100.0
4	43	75	2	0.97	0.96	0.82	0.81	0.89	85.3	14.7	100.0
4	44	74	2	0.22	0.23	0.41	0.45	0.53	82.4	17.6	100.0
4	96	79	2	1.00	1.08	0.72	0.76	0.53	60.8	35.4	96.2
4	97	79	3	1.05	1.01	1.06	1.08	0.94	86.1	13.9	100.0
4	99	79	2	0.65	0.62	0.70	0.67	0.81	82.3	17.7	100.0
4	107	79	2	1.35	1.29	0.83	0.83	0.88	87.3	11.4	98.7
4	115	77	3	1.40	1.40	0.96	1.00	0.95	89.6	10.4	100.0

Table D-6. Rater Reliability Statistics – Spanish Writing

Grade	Item	N	Max Score	Mean		SD		Weighted Kappa	Agreement		
				Rater 1	Rater 2	Rater 1	Rater 2		Exact	Adjacent	Exact + Adjacent
3	4	314	1	0.74	0.74	0.44	0.44	0.96	98.4	1.6	100.0
3	4	314	1	0.25	0.24	0.43	0.43	0.92	97.1	2.9	100.0
3	4	313	1	0.63	0.63	0.48	0.48	0.95	97.4	2.6	100.0
3	4	313	1	0.58	0.58	0.49	0.49	0.95	97.8	2.2	100.0
3	4	312	1	0.47	0.47	0.50	0.50	0.95	97.8	2.2	100.0
3	4	313	1	0.74	0.74	0.44	0.44	0.93	97.1	2.9	100.0
3	16	313	2	1.24	1.24	0.66	0.67	0.71	74.1	25.9	100.0
3	16	313	1	0.73	0.66	0.45	0.47	0.51	79.2	20.8	100.0
3	16	313	1	0.61	0.60	0.49	0.49	0.67	84.0	16.0	100.0
3	16	313	1	0.85	0.86	0.35	0.35	0.60	90.1	9.9	100.0
3	27	313	2	0.91	0.94	0.70	0.72	0.64	73.5	23.3	96.8
3	27	313	1	0.57	0.53	0.50	0.50	0.52	76.4	23.6	100.0
3	27	313	1	0.50	0.51	0.50	0.50	0.67	83.4	16.6	100.0
3	27	313	1	0.73	0.76	0.44	0.43	0.63	86.3	13.7	100.0
3	39	314	2	1.02	1.06	0.69	0.70	0.55	65.6	31.5	97.1
3	39	314	1	0.55	0.51	0.50	0.50	0.52	76.1	23.9	100.0
3	39	314	1	0.50	0.46	0.50	0.50	0.61	80.6	19.4	100.0
3	39	314	1	0.81	0.82	0.39	0.38	0.55	86.6	13.4	100.0
4	1	78	1	0.91	0.87	0.29	0.34	0.80	96.2	3.9	100.0
4	2	77	1	0.84	0.81	0.37	0.40	0.78	93.5	6.5	100.0
4	2	77	1	0.60	0.58	0.49	0.50	0.97	98.7	1.3	100.0
4	2	77	1	0.64	0.68	0.48	0.47	0.86	93.5	6.5	100.0
4	2	77	1	0.74	0.74	0.44	0.44	0.86	94.8	5.2	100.0
4	2	77	1	0.70	0.69	0.46	0.47	0.91	96.1	3.9	100.0
4	2	77	1	0.19	0.21	0.40	0.41	0.72	90.9	9.1	100.0
4	3	78	4	1.71	1.53	1.06	1.03	0.81	70.5	25.6	96.2
4	3	78	4	1.64	1.40	1.06	1.04	0.75	60.3	34.6	94.9
4	3	78	2	1.00	0.83	0.66	0.71	0.69	70.5	29.5	100.0
4	24	76	4	1.42	1.51	0.98	0.99	0.72	65.8	27.6	93.4
4	62	78	4	0.35	0.32	0.92	0.85	0.70	91.0	3.9	94.9
4	89	76	4	0.55	0.63	1.04	0.89	0.69	72.4	19.7	92.1

Table D-7. Rater Severity Study Results

Content	Grade	2004 Item Number	2006 Item Number	Valid N	Mean 2004	Mean 2006	Mean Difference	STD 2004	STD 2006	% Adjacent	% Exact	% Discrepant	Corr	Intra Class Corr	Weighted Kappa
MA	5	13	13	1053	2.38	2.43	-0.05	0.81	0.82	11.87	87.18	0.95	0.88	0.94	0.88
MA	5	15	15	1060	0.92	0.89	0.03	0.96	0.96	4.34	95.09	0.57	0.96	0.98	0.96
MA	5	46	46	1058	2.85	2.89	-0.04	1.24	1.15	22.68	75.52	1.80	0.90	0.95	0.89
MA	6	7	7	1053	2.38	2.43	-0.05	0.81	0.82	11.87	87.18	0.95	0.88	0.94	0.88
MA	6	12	11	1053	0.45	0.51	-0.06	0.73	0.75	6.74	92.78	0.47	0.93	0.96	0.92
MA	6	20	20	1065	2.10	2.08	0.02	1.03	1.06	15.02	84.79	0.19	0.93	0.96	0.93
MA	7	20	20	995	1.33	1.38	-0.05	1.53	1.53	20.00	77.69	2.31	0.93	0.97	0.93
MA	7	49	52	1013	1.37	1.21	0.16	0.83	0.78	27.25	68.61	4.15	0.68	0.83	0.67
MA	7	4	3	1053	0.45	0.51	-0.06	0.73	0.75	6.74	92.78	0.47	0.93	0.96	0.92
MA	7	48	48	1033	0.80	0.81	-0.01	0.91	0.90	15.00	84.03	0.97	0.88	0.94	0.88
MA	7	60	60	1012	2.01	2.01	0.00	1.45	1.47	25.20	70.55	4.25	0.90	0.95	0.90
MA	8	20	20	995	1.33	1.38	-0.05	1.53	1.53	20.00	77.69	2.31	0.93	0.97	0.93
MA	8	11	8	1013	1.37	1.21	0.16	0.83	0.78	27.25	68.61	4.15	0.68	0.83	0.67
MA	8	46	46	991	1.20	1.24	-0.04	1.19	1.19	14.53	84.76	0.71	0.94	0.97	0.94
MA	8	31	31	909	0.65	0.68	-0.03	0.68	0.69	4.84	95.05	0.11	0.95	0.97	0.94
MA	9	40	40	869	0.38	0.35	0.03	0.69	0.68	14.04	85.85	0.12	0.85	0.92	0.85
MA	9	14	15	881	0.36	0.34	0.02	0.67	0.73	12.26	86.83	0.91	0.84	0.92	0.84
MA	9	22	24	909	0.65	0.68	-0.03	0.68	0.69	4.84	95.05	0.11	0.95	0.97	0.94
MA	10	15	18	881	0.36	0.34	0.02	0.67	0.73	12.26	86.83	0.91	0.84	0.92	0.84
MA	10	6	4	922	1.14	1.20	-0.06	0.92	0.91	11.71	86.55	1.74	0.89	0.94	0.89
MA	10	60	60	862	1.01	1.09	-0.08	0.83	0.92	21.93	75.64	2.44	0.79	0.89	0.78

Table D-7. Rater Severity Study Results (continued)

Content	Grade	2004 Item Number	2006 Item Number	Valid N	Mean 2004	Mean 2006	Mean Difference	STD 2004	STD 2006	% Adjacent	% Exact	% Discrepant	Corr	Intra Class Corr	Weighted Kappa
RD	3	18	27	1012	1.02	1.00	0.02	0.84	0.84	8.79	90.71	0.49	0.92	0.96	0.92
RD	4	10	4	1083	1.53	1.48	0.05	1.28	1.27	6.28	86.98	6.74	0.89	0.94	0.89
RD	4	42	17	1068	2.01	2.06	-0.05	1.22	1.20	29.78	64.89	5.34	0.82	0.91	0.82
RD	4	105	109	1076	0.82	0.75	0.07	0.77	0.74	17.01	82.25	0.74	0.83	0.91	0.83
RD	5	38	25	1068	2.01	2.06	-0.05	1.22	1.20	29.78	64.89	5.34	0.82	0.91	0.82
RD	5	47	49	1066	1.25	1.27	-0.02	1.02	0.97	28.61	66.79	4.60	0.76	0.88	0.76
RD	6	11	13	1056	0.99	0.91	0.08	0.88	0.90	17.23	81.44	1.33	0.86	0.93	0.86
RD	6	53	45	1038	1.63	1.63	0.00	0.72	0.71	2.41	97.59	0.00	0.98	0.99	0.98
RD	7	10	10	1041	1.01	0.95	0.06	0.64	0.80	35.83	62.44	1.73	0.61	0.80	0.60
RD	7	100	103	1023	1.01	1.03	-0.02	0.94	0.94	24.54	73.31	2.15	0.81	0.90	0.81
RD	8	24	22	996	1.34	1.27	0.07	0.86	0.90	11.04	87.35	1.61	0.89	0.94	0.89
RD	8	113	114	904	1.69	1.70	-0.01	1.33	1.33	7.19	91.26	1.55	0.96	0.98	0.96
RD	9	29	26	987	2.24	2.13	0.11	1.04	1.11	25.03	70.62	4.36	0.81	0.90	0.81
RD	9	46	44	969	2.10	2.12	-0.02	0.90	0.86	29.00	68.01	2.99	0.73	0.87	0.73
RD	10	8	8	1052	1.22	0.83	0.39	0.64	0.71	46.10	50.00	3.90	0.50	0.69	0.43
RD	10	31	31	1023	0.88	1.03	-0.15	1.00	0.95	35.87	60.80	3.32	0.75	0.87	0.74
WR	3	39A	39A	1084	1.03	1.11	-0.08	0.38	0.40	16.42	82.84	0.74	0.39	0.69	0.38
WR	3	39B	39B	1084	0.93	0.96	-0.03	0.25	0.19	6.46	93.54	0.00	0.36	0.67	0.34
WR	3	39C	39C	1084	0.84	0.90	-0.06	0.36	0.30	14.11	85.89	0.00	0.39	0.69	0.37
WR	3	39D	39D	1084	0.90	0.93	-0.03	0.30	0.26	10.15	89.85	0.00	0.37	0.68	0.36
WR	4	52	53	1072	2.25	2.39	-0.14	0.89	0.80	28.17	67.72	4.10	0.68	0.83	0.67
WR	5	116	117	1082	2.38	2.34	0.04	0.66	0.69	24.21	74.95	0.83	0.70	0.85	0.70
WR	6	98	97	1045	2.24	2.23	0.01	0.68	0.72	28.13	71.39	0.48	0.70	0.85	0.69
WR	7	118	118	1028	2.32	2.05	0.27	0.72	0.65	34.44	65.08	0.49	0.70	0.81	0.64
WR	8	94	86	1021	2.28	2.24	0.04	0.76	0.57	30.17	69.64	0.20	0.69	0.83	0.66
WR	9	72	70	1037	2.26	2.34	-0.08	0.79	0.75	27.10	72.90	0.00	0.78	0.89	0.77
WR	10	76	73	1049	2.64	2.43	0.21	0.65	0.67	29.17	70.73	0.10	0.71	0.83	0.67

Table D-7. Rater Severity Study Results (concluded)

Content	Grade	2004 Item Number	2006 Item Number	Valid N	Mean 2004	Mean 2006	Mean Difference	STD 2004	STD 2006	% Adjacent	% Exact	% Discrepant	Corr	Intra Class Corr	Weighted Kappa
SC	8	70	70	1000	0.15	0.18	-0.03	0.35	0.38	5.50	94.50	0.00	0.81	0.90	0.80
SC	8	7	9	1015	0.85	0.83	0.02	0.80	0.78	9.75	90.15	0.10	0.92	0.96	0.92
SC	8	56	56	1024	1.00	0.98	0.02	0.83	0.81	18.46	80.86	0.68	0.84	0.92	0.84
SP RD	3	17	17	258	1.68	1.69	-0.01	0.63	0.63	12.40	87.21	0.39	0.82	0.91	0.82
SP RD	3	2	2	260	2.17	2.18	-0.01	1.20	1.19	3.46	96.54	0.00	0.99	0.99	0.99
SP RD	4	14	14	267	0.50	0.53	-0.03	0.56	0.56	19.48	80.15	0.37	0.67	0.83	0.67
SP RD	4	44	44	252	0.51	0.50	0.01	0.51	0.50	19.05	80.95	0.00	0.63	0.81	0.62
SP WR	3	27A	27A	291	1.03	0.83	0.20	0.69	0.66	26.12	70.45	3.44	0.61	0.79	0.58
SP WR	3	27B	27B	292	0.52	0.55	-0.03	0.50	0.50	32.88	67.12	0.00	0.34	0.67	0.34
SP WR	3	27C	27C	292	0.61	0.47	0.14	0.49	0.50	25.68	74.32	0.00	0.51	0.74	0.49
SP WR	3	27D	27D	292	0.83	0.69	0.14	0.38	0.46	18.84	81.16	0.00	0.54	0.74	0.50
SP WR	4	89	89	269	1.17	0.75	0.42	1.33	1.07	27.88	52.04	20.07	0.59	0.76	0.54