

**OCTOBER 2009 GRADE 3 READING TEST ADMINISTRATION
STATISTICAL SUMMARY**

These statistics describe the entire population of Ohio grade 3 public school students (including community schools) tested during the October 2009 administration.

Table 1 – Summary of Parameters from the October 2009 Administration of the Ohio Achievement Tests for Grade 3.

N-count	129,520
Max Raw Score	49
Raw Score Mean	31.54
Raw Score Standard Deviation	10.46
Raw Score SEM	3.41
Max Scaled Score	497
Scaled Score Mean	399.44
Scaled Score Standard Deviation	28.78
Scaled Score SEM	9.37
Reliability	0.89

Table 2 – Cut Score Points for Basic, Proficient, Accelerated, and Advanced Standards and Percentage of Students by Performance Levels

Standard	Raw Score	Scaled Score	Percent
Limited	Below 26	Below 385	28.50
Basic	26	385	18.21
Proficient	33	400	21.41
Accelerated	39	415	16.62
Advanced	43	432	15.27

Table 3 – Subscale (Raw Score) Bands for the Minimally Proficient Student in Grade 3, October 2009 Administration.

	Content Standard	Below	At or Near Proficient	Above	Possible Score
Grade 3 Reading	Vocabulary	0 – 7	8 – 9	10	10
	Reading process	0 – 9	10 – 13	14 – 16	16
	Informational text	0 – 7	8 – 10	11 – 12	12
	Literary text	0 – 6	7 – 9	10 – 11	11

Equating and Scaling: How Raw Scores Are Converted Into Scaled Scores

Test Form Construction

All items in the October 2009 Ohio Grade 3 Reading Achievement Test were previously field-tested, and item difficulty estimates from the field test administrations were used to pre-equate operational test forms during form construction. Pre-equating based on field-test item parameters aids in the development of nearly equivalent Ohio Achievement Test (OAT) forms across test administrations.

Common Item Equating

Following administration of the October 2009 Grade 3 Reading Achievement Test, item difficulty values were re-estimated using an early return sample. The early return sample was selected to be statistically representative of all Ohio grade 3 public school students. Because bank item difficulty estimates for all test items were available from the previous field test administrations, all the operational items could potentially serve as anchor items in the equating process. After calibrating items based on the early return sample data, ODE used a stepwise deletion procedure to calculate the linking constant needed to bring the set of operational items back to the OAT grade 3 reading “bank,” or reference, scale values. First, the October 2009 difficulty values (from the early return sample) were computed and compared with the bank difficulty values. The mean difference between the current and the bank difficulties of the anchor items is called the equating constant. The equating constant was added to each difficulty value for the current test administration item parameters so that the mean item difficulties are equal. Then the “linked” current item difficulty values were compared with the original bank values to identify the item with the largest absolute difference between the two values. If the absolute value of the difference was greater than 0.3 logits, the item was eliminated as an anchor item. This procedure was repeated until the largest difference between a linked current value and bank value was less than 0.3. This procedure ensures that the items used to anchor the operational test to the reference scale are stable. When the equating process is complete, item difficulties from the current administration are directly comparable with those from the bank.

Scaling

Because the meaning of raw scores changes across test forms and test administrations, scaled scores are usually used in place of raw scores.

As previously noted, after administering the October 2009 operational test, test items were calibrated and equated on the basis of the early return sample, and Rasch ability estimates were computed for each possible raw score. The Rasch ability estimates were then transformed to the Ohio Grade 3 Reading Achievement Test scale, which is scaled so that the Proficient standard is equal to 400.

Ohio Rounding Rule

When transforming raw scores to scaled scores, if the scaled score nearest to a performance standard cut score is below the cut score, then the scaled score is rounded up to equal the cut score. Otherwise, no special rounding is done. For example, if a raw score is associated with an observed scaled score of 398.94, and 398.94 is the closest observed scaled score to the Proficient level cut score, then this value is rounded up to 400, corresponding to the Proficient level performance standard. Conversely, if the closest scaled score value to the Proficient level cut score is 401.12, no special rounding rules are invoked, because the value is greater than the cut score.

Table 4 – Raw Score to Scaled Score Conversion Table

Raw Score	Scaled Score	Raw Score	Scaled Score
0	259	25	382
1	278	26	385
2	296	27	386
3	307	28	388
4	315	29	390
5	322	30	392
6	328	31	395
7	332	32	397
8	337	33	400
9	341	34	401
10	344	35	404
11	348	36	406
12	351	37	409
13	354	38	412
14	356	39	415
15	359	40	418
16	362	41	422
17	364	42	426
18	367	43	432
19	369	44	435
20	371	45	442
21	373	46	450
22	376	47	460
23	378	48	478
24	380	49	497