



**Developmental Scale for North Carolina
End-of-Grade/End-of-Course
ELA/Reading and English II Tests,
Fourth Edition**

**Alan Nicewander, Ph.D.
Tia Sukin, Ed.D.
Josh Goodman, Ph.D.
Huey Dodson, B.S.
Matthew Schulz, Ph.D.
Susan Lottridge, Ph.D.
Phoebe Winter, Ph.D.**

**Submitted to the
North Carolina Department of Education**

December 2, 2013

Pacific Metrics Corporation
1 Lower Ragsdale Drive
Building 1, Suite 150
Monterey, CA 93940

Developmental Scale for North Carolina End-of-Grade/End-of-Course ELA/Reading and English II Tests, Fourth Edition

This technical report describes the methods used and results found by Pacific Metrics Corporation in deriving the developmental scale for the North Carolina End-of-Grade/End-of-Course ELA/Reading and English II Tests, Fourth Edition. To create the vertical scale, Pacific Metrics used the methods already in place by North Carolina Department of Public Instruction (NCDPI) as described in the *North Carolina Reading Comprehension Tests Technical Report* (Bazemore & Van Dyke, 2004). For the ELA/Reading and English II scale, Pacific Metrics used Appendix C (Thissen, Edwards, Coon, & Woods, 2004) of that report. The article by Williams, Pommerich, and Thissen (1998) was also used as a reference.

Grade levels included in the Fourth Edition developmental scale slightly differ from those included in the First through Third editions. While First through Third edition scales include grades Pre 3 through 8, the Fourth Edition scale includes grades 3 through 8. The corresponding End-of-Course assessment, English II, was also included in the initial scale, but was dropped due to a North Carolina team decision.

Fourth Edition Developmental Scale

Table 1 presents the Fourth Edition developmental scale for the population for ELA/Reading and English II. Grade 5 was the base grade for the developmental scale, using a mean of 450 and standard deviation of 10. To create the developmental scale, the same items (called a linking set) were administered to students in adjacent grades. Both above- and below-grade links were used for the ELA/Reading and English II scale. Items were operational when on-grade level but served as embedded (e.g., did not contribute toward student scores) when placed off-grade level.

Table 1. Developmental Scale Means and Standard Deviations
Derived from Spring 2013 Item Calibration for
North Carolina End-of-Grade/End-of-Course Tests of Reading
Comprehension/English II, Fourth Edition

Grade	Population	
	Mean	Standard Deviation
3	440.01	10.90
4	446.00	10.33
5	450.00	10.00
6	452.70	10.99
7	455.97	11.12
8	458.66	11.35
English II	461.82	11.75

As shown in table 1 and as expected, the mean scores increased between grades, with growth ranging from 3 to 6 scale score points. The smallest increase occurred between grade 6 and grade 7; the largest increase occurred between grade 3 and grade 4.

The values for the developmental scales are based upon item response theory (IRT) estimates of differences between adjacent-grade mean thetas (θ) and ratios of adjacent-grade standard deviations of θ . The three-parameter logistic model was used to estimate item and person parameters. flexMIRT™ version 1.88 (Cai, 2012) was used. In flexMIRT™, the below grade was considered the reference group; its population mean and standard deviation were set to 0 and 1, respectively. The above-grade mean and standard deviation were estimated using the scored data and the IRT parameter estimates. These parameters were provided in the flexMIRT™ output and did not require independent calculation.

Individual runs in flexMIRT™ were conducted for each of the grade-pair links. For ELA/Reading, each grade pair for grades 3 through 8 had twelve links (six below-grade and six above-grade), and grade-pair 8–English II had thirty links (fifteen below-grade and fifteen above-grade). The linking sets varied between six and eight items, and each linking set was associated with a reading passage.

Under the assumption of equivalent groups, the form results were averaged within grade pairs to produce one set of values per adjacent grade. Outlying values were dropped if they were greater than two standard deviations from the mean. For ELA/Reading, three sets of values were dropped as outliers—one each from the 3–4, 6–7, and 7–8 grade pairs. Table 2 displays the average difference in adjacent-grade means and standard deviation ratios for Reading. Table 3 presents the mean difference and standard deviation ratio for each adjacent-grade link for Reading.

Table 2. Average Mean Difference in Standard Deviation Units of Lower Grade and Average Standard Deviation Ratios Derived from Spring 2013 Item Calibrations for North Carolina End-of-Grade/End-of-Course Tests of ELA/Reading and English II, Fourth Edition

Grades	Average Mean Difference	Average Standard Deviation Ratio	Number of Grade-Pair Forms
3–4*	0.550	0.948	11
4–5	0.387	0.968	12
5–6	0.270	1.099	12
6–7*	0.298	1.011	11
7–8*	0.242	1.021	11
8–English II	0.278	1.035	30

Note: An asterisk (*) denotes that one outlier was removed from the average for this grade pair.

Table 3. Values for Adjacent-grade Means in Standard Deviation (SD) Units of Lower Grade and Standard Deviation Ratios, Derived from Spring 2013 Item Calibrations for North Carolina End-of-Grade/End-of-Course Tests of ELA/Reading and English II, Fourth Edition

Grades 3–4		Grades 4–5		Grades 5–6		Grades 6–7		Grades 7–8		Grade 8–Eng II	
Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0.375	1.103	0.318	1.271	0.070	1.185	0.203	1.083	0.084	1.219	0.444	1.265
0.596	0.859	0.388	0.717	0.386	1.002	0.426	0.884	0.211	1.037	0.354	1.017
0.515	0.906	0.403	0.937	0.188	1.096	0.319	1.102	0.231	1.030	0.107	1.375
0.608	1.130	0.427	1.000	0.262	1.119	0.266	1.064	0.155	1.310	0.460	1.002
0.480	1.065	0.294	0.887	0.235	1.350	0.243	1.116	0.155	1.043	0.532	0.853
0.519	0.928	0.365	0.919	0.282	0.974	0.289	0.805	0.328	1.005	0.269	1.020
0.682	0.774	0.535	0.755	0.421	0.858	0.391	0.720	0.303	0.797	0.583	0.922
0.588	0.950	0.498	0.987	0.355	0.953	0.411	1.021	0.193	1.113	0.643	0.696
0.561	0.908	0.308	1.095	0.300	1.160	0.257	1.040	0.363	0.995	-0.036	1.429
0.533	0.878	0.457	0.831	0.329	1.117	0.323	0.912	0.376	1.005	-0.133	1.245
0.506	1.016	0.346	1.038	0.303	1.153	0.277	1.043	0.264	0.949	0.400	1.163
0.465	1.014	0.302	1.176	0.103	1.221	0.268	1.056	0.151	1.034	0.292	0.868
										0.383	1.019
										0.310	0.968
										-0.025	1.346
										0.477	0.829
										0.441	0.822
										0.090	0.846
										0.426	0.953
										0.552	0.726
										-0.150	1.368
										-0.093	1.227
										0.182	1.229
										0.268	0.818
										0.227	1.070
										0.487	0.866
										0.216	1.086
										0.411	0.788
										0.119	1.277
										0.115	0.969

Note: Means and standard deviations in shaded cells were dropped from analyses as outliers.

Comparison of Fourth Edition Developmental Scale to First through Third Edition Scales

Table 4 presents the mean scale scores by grade for the First, Second, Third, and Fourth editions for ELA/Reading and English II. To facilitate comparison of the growth between grades among the First, Second, Third, and Fourth editions, figure 1 presents the mean scores plotted together for ELA/Reading and English II. To place the First, Second, Third, and Fourth edition scores on similar scales, a value of 300 was added to the First Edition scores, a value of 200 was added to the Second Edition scores, and a value of 100 was added to the Third Edition scores.

For ELA/Reading and English II, greater average growth between grades 3–8 occurred in the Third Edition (19.72) than in the First, Second, and Fourth editions (13.96, 14.14, and 18.65, respectively). As shown in figure 1, the First through Fourth editions exhibited similar growth in mean scores between grades 3–8.

Table 4. Comparison of Population Means and Standard Deviations for First through Fourth Editions of North Carolina End-of-Grade/End-of-Course Tests of ELA/Reading and English II

Grade	First Edition (1992)		Second Edition (2002)		Third Edition (2008)		Fourth Edition (2013)	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Pre 3	139.02	8.00	236.66	11.03	326.56	14.64		
3	145.59	9.62	245.21	10.15	338.62	12.57	440.01	10.90
4	149.98	9.50	250.00	10.00	345.20	10.79	446.00	10.33
5	154.74	8.21	253.92	9.61	350.00	10.00	450.00	10.00
6	154.08	9.44	255.57	10.41	352.86	10.12	452.70	10.99
7	157.81	9.09	256.74	10.96	355.63	9.79	455.97	11.12
8	159.55	8.96	259.35	11.13	358.34	9.49	458.66	11.35
English II							461.82	11.75

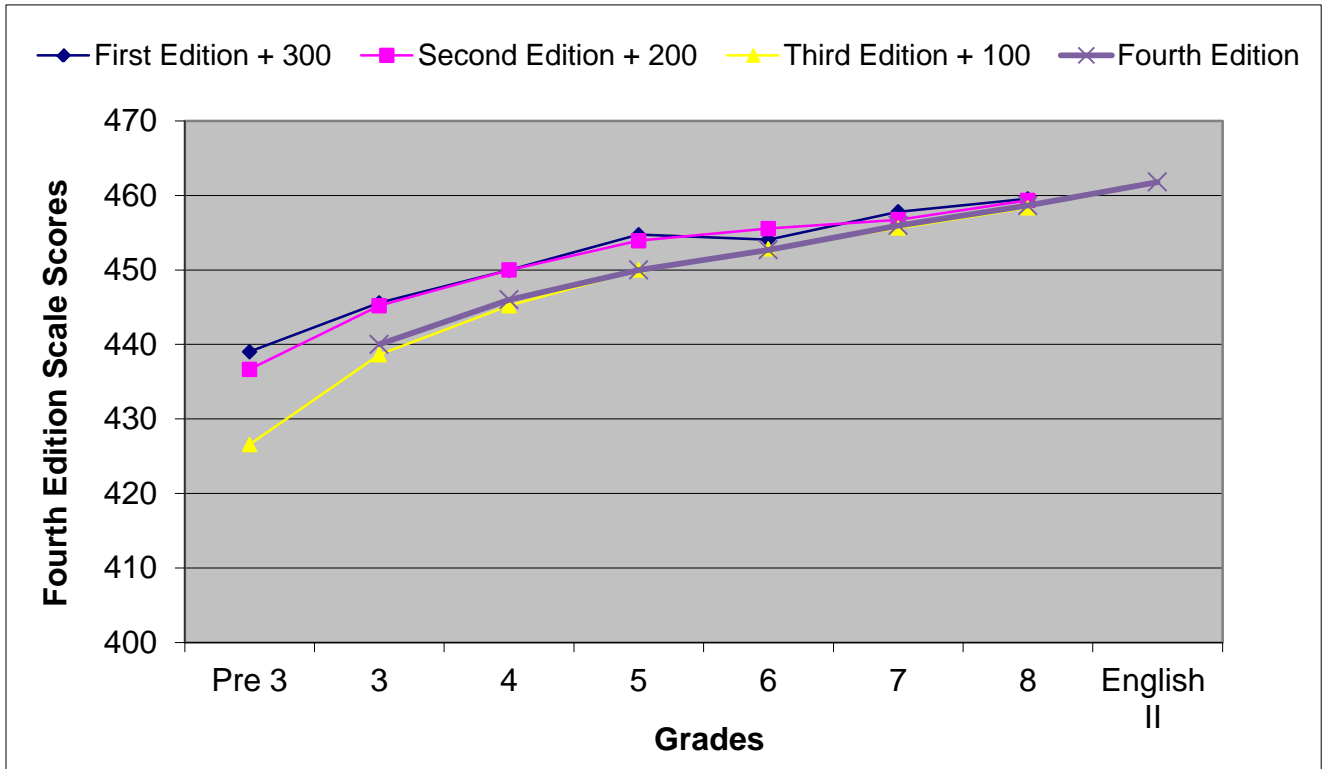


Figure 1. Comparison of Growth Curves between First, Second, Third, and Fourth Editions of North Carolina End-of-Grade/End-of-Course Tests of ELA/Reading and English II.

Quality Assurance Procedures

The authors have applied a variety of analyses and procedures to ensure that the results of the scaling and linking studies are correct. For the vertical scale, the mean difference and standard deviation ratios for the grades and forms were compared to the classical test theory p -values of the linking items. The data provided evidence that the mean difference and standard deviation ratios were accurate in both direction and magnitude (see table 5). Also, previous work using the described statistical method to create the vertical scale was applied to the Second Edition data to ensure that it reproduced the scale correctly.

Table 5. Average Mean Difference in Standard Deviation Units of Lower Grade and Standard Deviation Ratios, and Average Difference in p -values (Higher Minus Lower Grade) of Linking Sets, for North Carolina End-of-Grade/End-of-Course Tests of ELA/Reading and English II, Fourth Edition

Grade Pair	Average Mean Difference	Mean p -value Difference for Linking Items
3–4*	0.550	0.097
4–5	0.387	0.068
5–6	0.270	0.044
6–7*	0.298	0.049
7–8*	0.242	0.046
8–English II	0.278	0.050

Note: An asterisk (*) denotes that one grade-pair link was dropped from analyses as an outlier.

Additionally, IRT parameters provided separately by the North Carolina Department of Education were correlated with the flexMIRT™ calibrated item parameters within grade pairs and averaged across grades. For Reading, the average correlation for discrimination parameters was 0.97 with a standard deviation of 0.01 across grade and form pairs. The average correlation for difficulty or step parameters (for English II multi-point items) was 0.97 with a standard deviation of 0.02. The average correlation for guessing parameters was 0.93 with a standard deviation of 0.02.

Psychometrics Underlying the Developmental Scale

The procedure for creating the developmental scale is based upon that described in Williams, Pommerich, and Thissen (1998). The procedure is divided into four steps, described below.

Step 1. flexMIRTTM was used to calibrate the End-of-Grade and End-of-Course Reading tests' item and population parameters for adjacent grades. This process was described in the section entitled "Fourth Edition Developmental Scale" of this report and resulted in average mean difference and average standard deviation ratios (m_n and s_n) for each grade n (see table 2).

Step 2. A (0,1) growth scale anchored at grade 3 was constructed to yield the following means (M_n) and standard deviations (S_n):

$$M_n = M_{n-1} + m_n S_{n-1}, \quad \text{mean for Grade } n \text{ on (0,1) growth scale anchored at the lowest grade (with grade 3 indexed as } n=3\text{),}$$

$$S_n = s_n S_{n-1}, \quad \text{standard deviation for grade } n \text{ on (0,1) growth scale anchored at the lowest grade (with grade 3 indexed as } n=3\text{),}$$

where $M_2 \equiv 0$, and $S_2 \equiv 1$. This (0,1) growth scale was generated recursively upwards to the End-of-Course (English II).

Step 3. The scale was re-centered (re-anchored) at grade 5, yielding

$$M_n^* = \frac{(M_n - M_5)}{S_5}$$

$$S_n^* = \frac{S_n}{S_5}$$

as the means (M_n^*) and standard deviations (S_n^*).

Step 4. The final step in constructing the growth scale was the application of a linear transformation in order to produce a growth scale with the grade 5 mean and standard deviations equal to 450 and 10, respectively, *viz.*,

$$\mu_n = 450 + 10 M_n^*$$

$$\sigma_n = 10 S_n^*,$$

where μ_n is the mean of the final growth scale in grade n and σ_n is the standard deviation for the growth scale in grade n .

References

- Bazemore, M., & Van Dyke, P. (2004). *North Carolina Reading Comprehension Tests Technical Report*. Raleigh, NC: North Carolina Department of Public Instruction.
- Cai, L. (2012). flexMIRT™ version 1.88: A numerical engine for multilevel item factor analysis and test scoring. [Computer software]. Seattle, WA: Vector Psychometric Group.
- Thissen, D., Edwards, M., Coon, C. & Woods, C. (2002). *North Carolina Reading Comprehension Tests Technical Report, Appendix C*. Raleigh, NC: North Carolina Department of Public Instruction.
- Williams, V.S.L., Pommerich, M., & Thissen, D. (1998). A comparison of developmental scales based upon Thurstone methods and item response theory. *Journal of Educational Measurement*, 35, 93-107.