



Research Memorandum
ETS RM-11-15

**An Examination of United States
Middle School Students'
Performance on TOEFL® *Junior*™**

Mikyung Kim Wolf

Jonathan Steinberg

September 2011

An Examination of United States Middle School Students' Performance on TOEFL® *Junior*™

Mikyung Kim Wolf and Jonathan Steinberg
ETS, Princeton, New Jersey

September 2011

As part of its nonprofit mission, ETS conducts and disseminates the results of research to advance quality and equity in education and assessment for the benefit of ETS's constituents and the field.

To obtain a PDF or a print copy of a report, please visit:

<http://www.ets.org/research/contact.html>

Technical Review Editor: Dan Eignor

Technical Reviewers: Rick Morgan and John Young

Copyright © 2011 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, LISTENING. LEARNING. LEADING., and TOEFL are registered trademarks of Educational Testing Service (ETS). TOEFL JUNIOR is a trademark of ETS.



Abstract

The purpose of this study was to examine how native English-speaking and English language learner (ELL) students in U.S. schools perform on the TOEFL[®] *Junior*[™] test. Specifically, this study was designed to provide percentile rank information of the TOEFL *Junior* scores for both native English-speaking and ELL students who attend U.S. schools. As the TOEFL *Junior* test is intended to measure English language skills representative of English-medium instructional environments at the middle-school level, this study focused on students in Grades 6 to 8. The percentile rank and statistical significance test results indicated that there was a clear pattern of increasing performance of the sample students on the TOEFL *Junior* test by grade level. It should be noted that the students in this sample are not truly representative of all middle school students and that interpretations should be made with caution. Suggestions for future studies are included in this report.

Key words: TOEFL *Junior*, percentile ranks, smoothed distributions, native English speakers, English language learners, middle-school grades

Acknowledgments

We would like to thank the following individuals for their guidance of this study and review of earlier drafts of this report: Brent Bridgeman, Rick Morgan, Don Powers, and Rick Tannenbaum. We also thank John Young and Dan Eignor for additional technical reviews. We are grateful to Rachel Adler, Wendy Albertson, Michaela Arzt, Jennifer Brown, and Peter Ciemins for their generous research support and assistance. We thank Ayleen Stelhorn and Kim Fryer for copyediting this report. Finally, we thank the participating schools and students for taking the TOEFL *Junior* test as part of this study.

Goal of the Study

This document reports results of a study that examined how native English-speaking and English language learner (ELL) students in U.S. schools performed on the TOEFL[®] *Junior*[™] test. The TOEFL *Junior* test is intended to measure the degree to which middle school students have attained proficiency in academic and social English-language skills, representative of English-medium instructional environments. The test consists of three sections: Listening Comprehension, Language Form and Meaning, and Reading Comprehension. The study was conducted to address test users' requests to provide information about TOEFL *Junior* scores in relation to the performance of middle school students in the United States where English is a medium of instruction.

Specifically, the goal of the study was to provide percentile rank information of the TOEFL *Junior* total scaled scores and corresponding section scaled scores for both native English-speaking and ELL students who attend U.S. schools. Considering that the TOEFL *Junior* test was intended for English learners worldwide, it was of particular interest to examine the performance of ELL students in U.S. schools on the test. This study focused on Grades 6 to 8. In the United States, Grade 6 may fall into either an elementary or a middle school grade while Grades 7 and 8 generally fall into middle school grades. Due to time and cost constraints, the sample was not truly representative of all middle schools. Thus, it should be noted that the percentile rank information may not represent, or generalize to, all middle school students.

Method

Sample

As noted above, it was not possible to obtain a representative sample of middle school students in the United States. In order to control, at least somewhat, for a wide range of variability of students, midperforming schools were targeted. The GreatSchools website (<http://www.greatschools.org/>), which rates U.S. schools based on academic test scores on a scale of 0 (*lowest*) to 10 (*highest*), was used as a resource. Midperforming schools with ratings of 4 to 6 were contacted. In total, 2,622 students in Grades 6 to 8 from 13 schools across seven states participated in the study. The states included California, Georgia, Kansas, New Hampshire, New Jersey, Texas, and Virginia. Of these students, 1,596 were native speakers of English, and 1,026 were ELLs. ELLs whose first language was not English were identified by each school primarily based on English-language proficiency tests. The group of students was

balanced in terms of gender, with a slight majority toward female students (51%). Table 1 summarizes the number of participating students by grade and by native English-speaking status.

Of the 13 participating schools, two were from the Northeast, five from the West, two from the Midwest, and four from the South. Over 60% of the students were from the South. The majority of the schools were from urban school districts, yielding 98% of the participating students from urban schools. In terms of the GreatSchools rating, nine schools were ranked at 6, three at 5, and one at 4. Tables 2 and 3 present the distributions of the sample students by region, school, and GreatSchools ranking.

Table 1

Frequency of Participating Students by Grade and Native English-Speaking Status

Grade	Native speakers of English	English language learners	Total
6	452	288	740
7	547	427	974
8	597	311	908
Total	1,596	1,026	2,622

Table 2

Frequency Distribution of Participating Students Across Regions and Native English-Speaking Status

Region	Schools	Native speakers of English	English language learners	Total
Northeast (NH, NJ)	2	76	36	112
West (CA)	5	364	397	761
Midwest (KS)	2	31	23	54
South (GA, TX, VA)	4	1,125	570	1,695
Total	13	1,596	1,026	2,622

Table 3

Frequency Distribution of Participating Students by School Rankings According to GreatSchools

GreatSchools rank ^a	Schools	Students
Rank 4	1	14
Rank 5	3	829
Rank 6	9	1,779
Total	13	2,622

^a The GreatSchools website (<http://www.greatschools.org/>) rates U.S. schools based on academic test scores on a scale of 0 (*lowest*) to 10 (*highest*).

A short questionnaire was developed to gather participating students' background information including students' academic performance indicated by states' standardized reading/English language arts (ELA) tests. For ELL students, information about their home language, grade at which they entered U.S. schools, and their levels of English proficiency as measured by states' standardized English-language proficiency tests was also collected. The questionnaire was completed by the teachers. Depending on the questions, all information was not obtained for all students.

State standardized reading/ELA tests are taken by all students, whereas English-language proficiency tests are taken only by ELL students. English-language proficiency tests typically contain four sections: listening, reading, speaking, and writing. Reading/ELA tests and reported English-language proficiency levels (e.g., Basic, Intermediate, Proficient, Advanced) varied across states. Teachers reported the students' proficiency levels.

Table 4 summarizes the participating students' academic performances on the states' standardized reading/ ELA tests. Approximately 70% of the native English-speaking students and 47% of ELL students in our sample reached at least the proficient level on their states' reading/ELA tests.

Table 4***Frequency Distribution of Students Who Met the Proficiency Standards on Their States' Standardized Reading/English Language Arts Tests***

Proficiency level	Native speakers of English	English language learners
Proficient and above	1,032 (70%)	441 (47%)
Below proficient	432 (30%)	503 (53%)
Total	1,464 (100%)	944 (100%)

Regarding the backgrounds of the participating ELL students, 78% of the ELL students in this study entered U.S. schools at Kindergarten or Grade 1. Among the 909 ELL students whose home language information was collected, Spanish was the most frequent home language spoken (86.9%). Other home language responses included Vietnamese (5.9%), Chinese (0.4%), Korean (0.4%), French (0.2%), and others (6.1%). Regarding the English-language proficiency tests (taken only by ELLs), proficiency levels were collected for only 59% of the ELL sample. On these tests, approximately 70% of the students reached at least the proficient level on their states' standardized English-language proficiency tests.

Procedure

The participating students took the October 2010 test form. In order to examine the data for reasonableness, descriptive statistics including means and standard deviations of scaled scores were first examined. Statistical significance tests were performed to examine whether the test scores differed by grade level. Specifically, multivariate analysis of variance (MANOVA) was conducted with the scaled scores from Listening Comprehension, Language Form and Meaning, Reading Comprehension, and total scaled scores as dependent variables. Pairwise post-hoc comparisons were also conducted to detect which grade levels performed differently on each section and the total test.

Percentile ranks were computed separately for native English-speaking and ELL students at each grade level. Due to the small sample sizes, score distributions were smoothed using a three-moment log-linear smoothing model (Moses & von Davier, 2006) that preserved the mean, standard deviation, and skewness of the underlying distribution. In the case of a small sample size, it is desirable to estimate the distribution of the population based on the observed sample and to use the estimated distribution to alleviate concerns about the instability of the small

sample. Smoothing the score distributions yet preserving the sample means and standard deviation is one technique of estimating the population distributions (Livingston, 1993). The percentile ranks were created based on the smoothed distribution of the observed sample. The technical details of the procedure used to create the percentile ranks presented in this report are available in a separate paper (see Steinberg & Moses, 2011).

Results

Descriptive Statistics

The score scales of the TOEFL *Junior* test ranged from 200 to 300 for each section and 600 to 900 for the total test. Table 5 and Table 6 present the descriptive statistics of the test scores for native English-speaking students and ELL students by each grade level, respectively. The native English-speaking and ELL students' total score distributions at each grade level is displayed in Figures A1 to A6 in the appendix.

As shown in these tables and the accompanying figures in the appendix, the students at higher grades performed better than those at lower grades. The mean difference in scaled scores among the students at different grade levels was statistically significant for each test section and for the total test. Table 7 summarizes the MANOVA test of differences in scaled scores across groups, along with the post-hoc comparison results. For example, the performances of native English-speaking students on the total test were statistically different across all three grade levels. On the other hand, ELL students' performances on the total test were statistically different between Grade 6 and Grades 7 and 8.

Table 5

Descriptive Statistics for Native Speakers of English in Grades 6 to 8

Section	Grade 6 (<i>N</i> = 452)				Grade 7 (<i>N</i> = 547)				Grade 8 (<i>N</i> = 597)			
	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Total	630	895	785	65.0	620	900	798	61.6	605	900	809	62.2
LC	210	300	278	18.5	205	300	280	17.6	200	300	284	17.6
LFM	200	300	260	22.6	200	300	266	21.9	200	300	268	22.3
RC	200	300	247	31.5	200	300	252	30.0	200	300	257	30.5

Note. Total = total scaled score, LC = listening comprehension scaled score, LFM = language form & meaning scaled score, RC = reading comprehension scaled score.

Table 6***Descriptive Statistics for English Language Learners in Grades 6 to 8***

Section	Grade 6 (N = 288)				Grade 7 (N = 427)				Grade 8 (N = 311)			
	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Total	630	885	743	61.1	605	895	768	62.5	620	885	777	62.7
LC	205	300	267	20.9	200	300	272	20.2	200	300	274	20.8
LFM	200	295	248	21.8	200	300	255	21.9	200	300	260	21.9
RC	200	295	229	27.2	200	295	240	28.7	200	295	243	29.4

Note. Total = total scaled score, LC = listening comprehension scaled score, LFM = language form & meaning scaled score, RC = reading comprehension scaled score.

Table 7***Summary of Multivariate Analysis of Variance (MANOVA) Results and Post-Hoc Comparisons***

Section	Native speakers of English		English language learners	
	Significance	Post-hoc comparison	Significance	Post-hoc comparison
Total	$F=11.8, df=2, p=.000$	Gr8 > Gr7 > Gr6	$F=11.3, df=2, p=.000$	Gr7, 8 > Gr6
LC	$F=19.9, df=2, p=.000$	Gr8 > Gr6, 7	$F=20.5, df=2, p=.000$	Gr8 > Gr7 > Gr6
LFM	$F=14.9, df=2, p=.000$	Gr7, 8 > Gr6	$F=22.1, df=2, p=.000$	Gr7, 8 > Gr6
RC	$F=19.3, df=2, p=.000$	Gr8 > Gr7 > Gr6	$F=23.4, df=2, p=.000$	Gr7, 8 > Gr6

Note. Total = total scaled score, LC = listening comprehension scaled score, LFM = language form & meaning scaled score, RC = reading comprehension scaled score.

Percentile Ranks

The percentile ranks for total scaled scores as well as scaled scores for each section for both native speakers of English and ELLs at each grade are displayed in Tables 8 to 11. Because only a small number of students obtained total scores at the lower end of the scale (e.g., 670 or less), the percentile ranks for the lower scaled scores may be less stable than those for the higher scores.

Table 8***Percentile Ranks for Total Scaled Scores for Native English Speakers***

Total scaled score	Grade 6 (N = 452)	Grade 7 (N = 547)	Grade 8 (N = 597)
900	99	99	99
890	97	96	95
880	94	93	90
870	90	88	84
860	86	83	78
850	82	78	71
840	77	72	64
830	72	66	57
820	67	59	50
810	62	53	43
800	56	47	37
790	50	41	32
780	45	35	27
770	39	30	23
760	34	25	19
750	29	21	16
740	25	17	13
730	21	14	11
720	17	12	9
710	14	9	8
700	11	7	7
690	9	6	6
680	7	5	5
670	5	3	4
660	4	3	3
650	3	2	3
640	2	1	2
630	1	1	2
620	1	1	1
610	0	0	1
600	0	0	0
Mean scaled score	785	798	809

Table 9***Percentile Ranks for Three Sections of Scaled Scores for Native English Speakers***

Scaled score	Listening Comprehension			Language Form & Meaning			Reading Comprehension		
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8
300	94	93	90	99	98	98	99	99	98
295	81	78	72	96	94	93	95	96	94
290	68	64	55	92	88	85	91	91	88
285	56	52	41	87	80	76	86	86	80
280	45	40	30	80	71	66	81	79	72
275	36	31	22	72	62	55	76	72	64
270	27	23	16	64	52	45	70	65	56
265	21	17	11	55	43	36	65	59	49
260	16	13	8	46	34	28	60	52	43
255	12	9	6	38	27	22	55	47	38
250	8	7	5	31	21	17	51	42	34
245	6	5	4	25	17	14	47	37	31
240	4	3	3	19	13	11	43	33	28
235	3	2	3	15	10	9	39	30	25
230	2	2	2	11	8	7	35	27	23
225	1	1	2	8	6	6	32	24	21
220	1	1	2	6	4	5	28	21	19
215	1	1	1	4	3	4	24	18	16
210	0	0	1	3	2	3	19	14	13
205	0	0	1	2	1	2	13	10	9
200	0	0	0	0	0	1	5	4	4
Mean scaled score	278	280	284	260	266	268	247	252	257

Table 10***Percentile Ranks for Total Scaled Scores for English Language Learners***

Total scaled score	Grade 6 (N=288)	Grade 7 (N=427)	Grade 8 (N=311)
900	99	99	99
890	99	99	99
880	98	98	97
870	97	96	95
860	96	94	92
850	94	91	88
840	93	88	84
830	90	83	79
820	88	78	73
810	85	73	67
800	82	67	61
790	78	61	55
780	73	55	48
770	68	48	42
760	63	42	37
750	57	37	31
740	51	31	26
730	44	26	22
720	38	22	18
710	31	18	15
700	25	15	12
690	20	12	10
680	15	10	8
670	11	8	6
660	8	6	5
650	5	4	4
640	3	3	3
630	2	2	2
620	1	1	1
610	0	1	1
600	0	0	0
Mean scaled score	743	768	777

Table 11***Percentile Ranks for Three Sections of Scaled Scores for English Language Learners***

Scaled score	Listening Comprehension			Language Form & Meaning			Reading Comprehension		
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8
300	98	97	95	99	99	99	99	99	99
295	93	90	85	99	99	98	99	99	98
290	86	80	75	98	97	95	98	98	96
285	79	70	64	96	93	90	97	96	93
280	70	59	53	93	88	83	95	92	89
275	61	48	43	89	81	74	92	88	84
270	52	38	35	83	73	65	90	82	78
265	43	30	28	76	63	55	86	75	72
260	35	23	21	69	54	45	83	69	65
255	28	17	17	60	45	36	79	62	60
250	21	13	13	51	36	29	76	56	54
245	16	10	10	43	29	23	72	51	49
240	12	7	7	35	23	18	68	47	45
235	8	6	6	28	18	14	64	43	41
230	6	4	4	21	14	11	60	39	37
225	4	3	3	16	11	8	55	36	33
220	3	3	2	12	8	6	50	33	30
215	2	2	2	8	6	5	44	29	26
210	1	1	1	5	4	3	36	24	21
205	0	1	1	3	2	2	26	17	15
200	0	0	0	1	1	1	10	7	5
Mean scaled score	267	272	274	248	255	260	229	240	243

As indicated in the descriptive statistics by grade, the percentile rank results demonstrated that students' performances increased by grade level. The percentile ranks indicated the percentage of examinees scoring lower than a given scaled score point. For example, as shown in Table 8, the total scaled score of 850 is at the 82nd percentile for Grade 6 native speakers of English who took this test, but at the 78th percentile for Grade 7 native speakers of English.

Concluding Remarks

The percentile rank information was based upon the sample of this study. As mentioned earlier, the sample group of this study is not representative of all U.S. middle school students. The relatively small sample size at the subgroup level should also be taken into consideration when interpreting the results. To increase the stability of the information provided here, a further study with a larger, national sample should be conducted. Yet, we still hope that the results of the study provide one piece of useful information about these groups' performances on the TOEFL *Junior* test. A promising result is that the TOEFL *Junior* test was able to distinguish the sample middle school students' performances by grade level. Another interesting note is that reading comprehension section scaled scores were more distinguishable across the grade levels than other section scaled scores for native English-speaking students. As for a future study, an investigation into students' performances at the item level will be valuable to understand students' performance in relation to the constructs measured by the TOEFL *Junior* test.

References

- Livingston, S. A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement*, 30(1), 23–39.
- Moses, T. P., & von Davier, A. A. (2006). *A SAS macro for loglinear smoothing: Applications and implications* (ETS Research Report No. RR-06-05). Princeton, NJ: ETS.
- Steinberg, J., & Moses, T. (2011, October). Smoothing scaled score distributions from a standardized test using PROC GENMOD. In *SESUG 2011: The 19th annual proceedings of the SouthEast SAS Users Group, Arlington, VA*.

Appendix
Graphs of Score Distribution by Grade

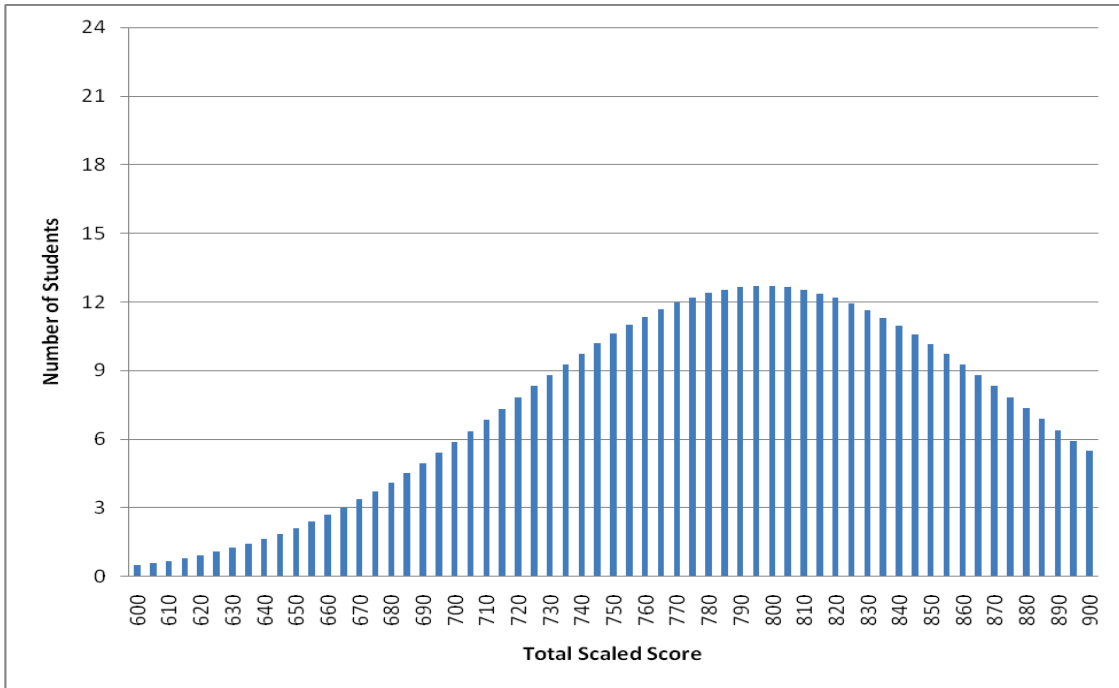


Figure A1. Total score distribution of Grade 6 native speakers of English ($N = 452$).

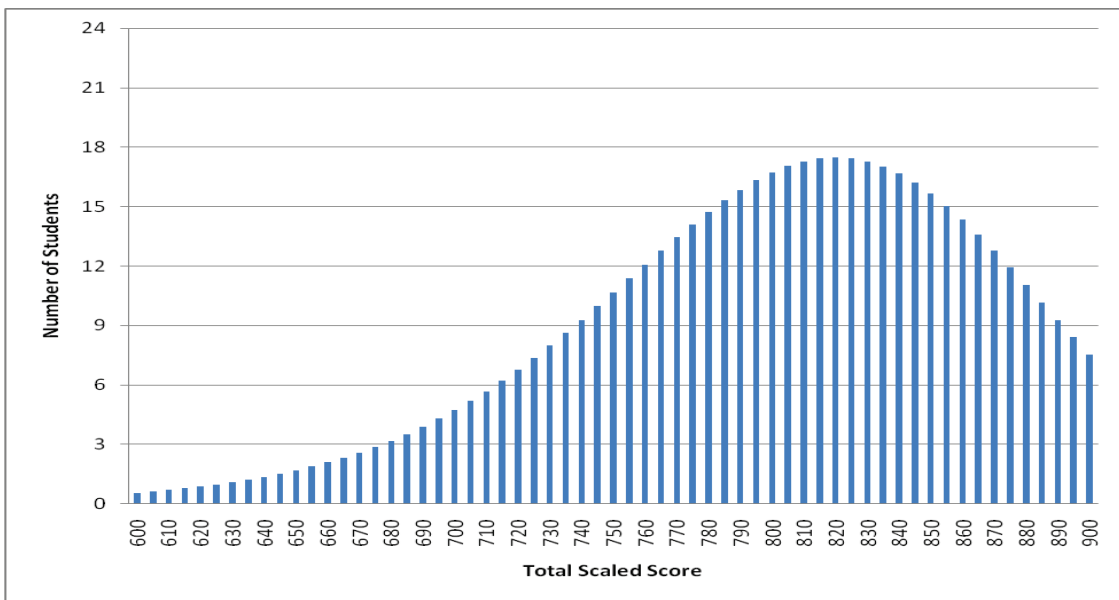


Figure A2. Total score distribution of Grade 7 native speakers of English ($N = 547$).

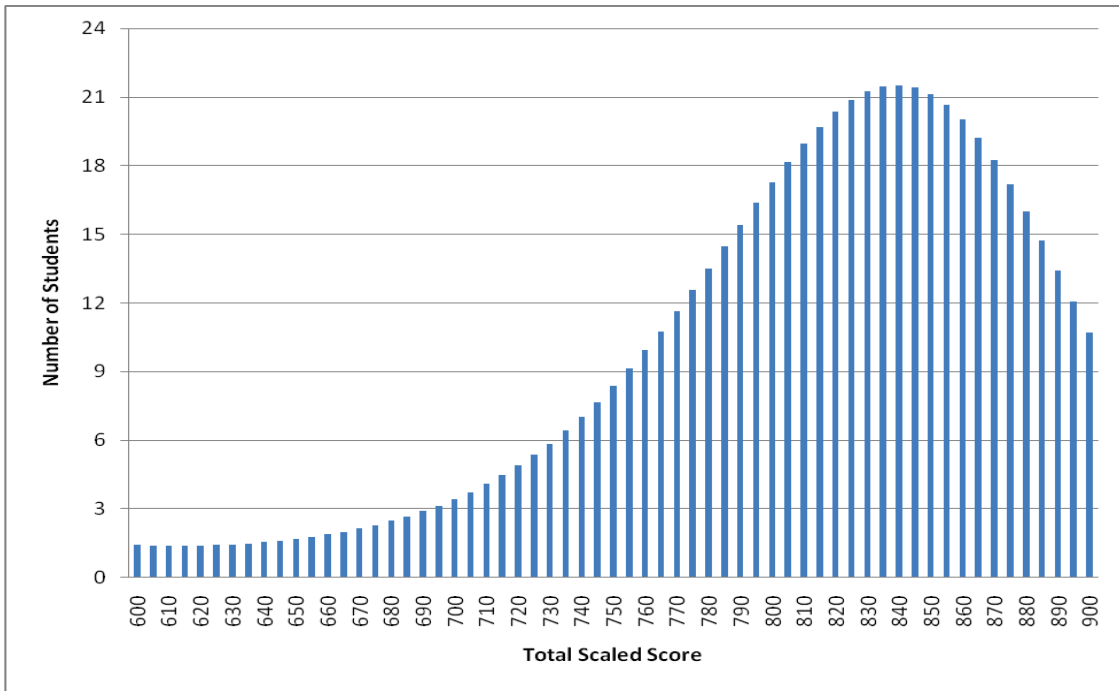


Figure A3. Total score distribution of Grade 8 native speakers of English ($N = 597$).

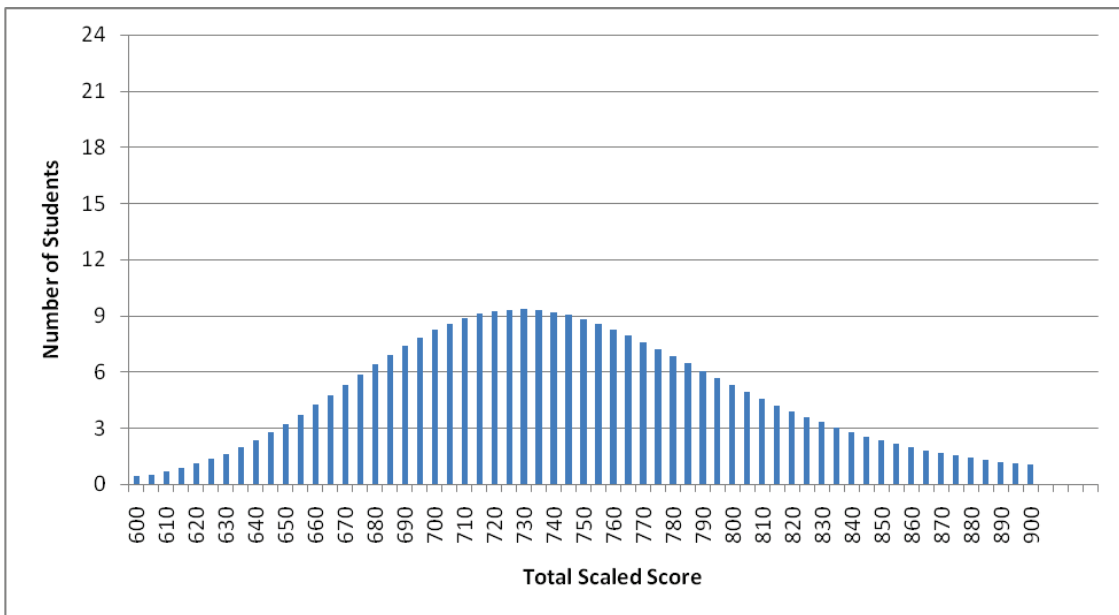


Figure A4. Total score distribution of Grade 6 English language learners ($N = 288$).

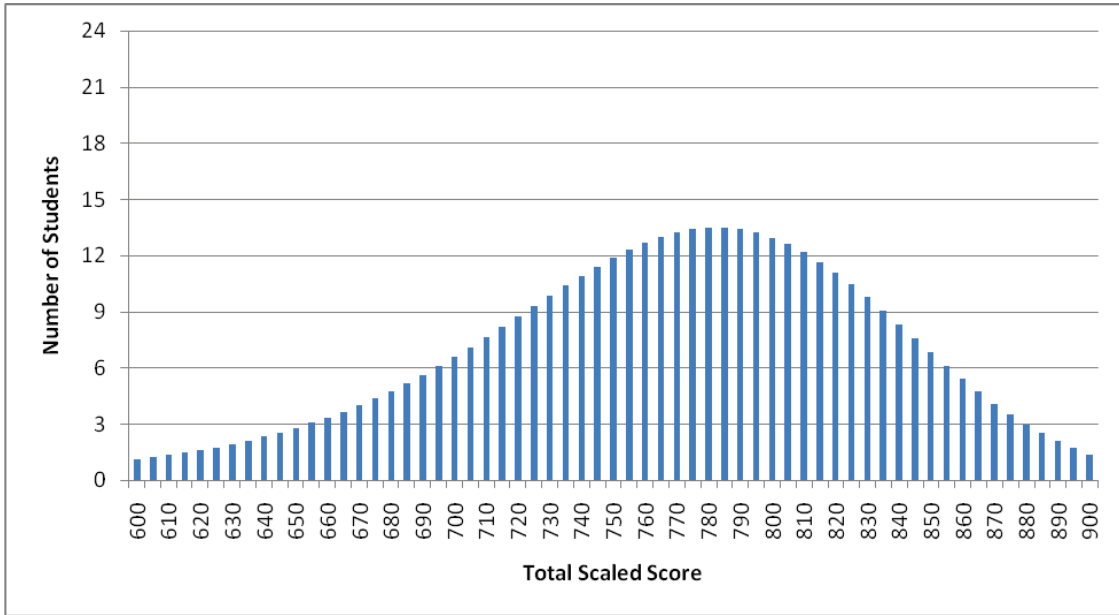


Figure A5. Total score distribution of Grade 7 English language learners (N = 427).

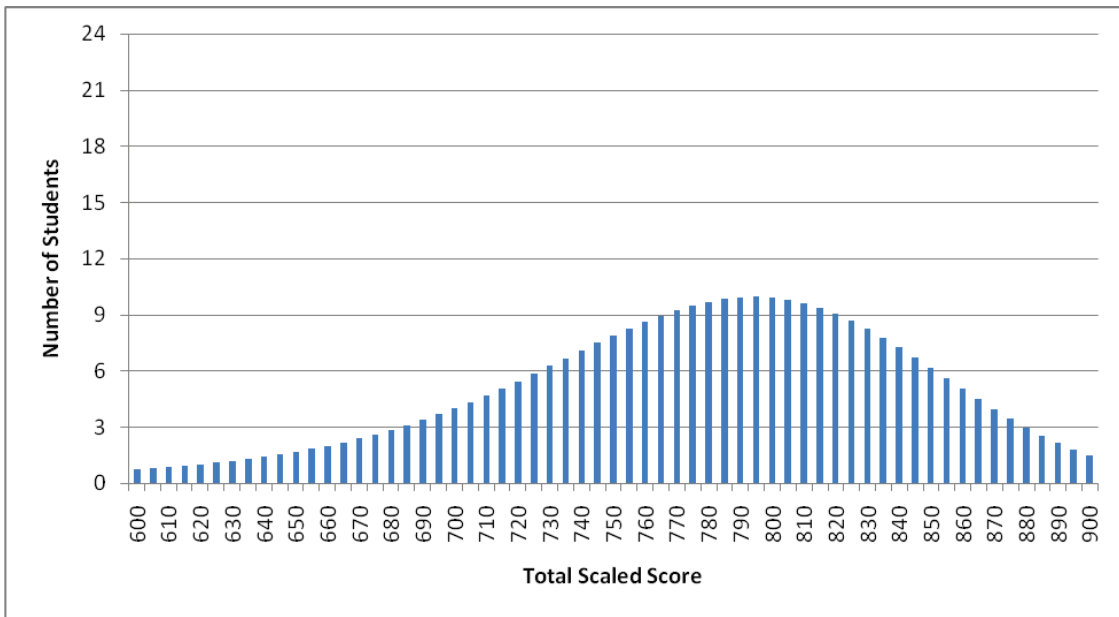


Figure A6. Total score distribution of Grade 8 English language learners (N = 311).