# OHIO TEST OF ENGLISH LANGUAGE ACQUISITION (OTELA) MARCH 2008 ADMINISTRATION

# STATISTICAL SUMMARY

These statistics describe the population of Ohio limited English proficient (LEP) students completing all domains in the March 2008 OTELA administration.

			OTELA Descriptive Statistics by Domain and Grade Cluster									
			Max	Raw	Raw Score	Raw	Max	Scaled	Scaled Score	Scaled		
Grade		N-	Raw	Score	Standard	Score	Scaled	Score	Standard	Score		
Cluster	Domain	count	Score	Mean	Deviation	SEM	Score	Mean	Deviation	SEM	Reliability	
	Listening	4116	21	13.23	5.42	1.33	432.00	341.34	56.47	13.83	0.94	
V	Speaking	4116	24	15.47	6.48	1.31	429.00	333.72	60.55	12.23	0.96	
К	Reading	4116	42	23.66	11.13	2.43	383.00	308.48	32.87	7.17	0.95	
	Writing	4116	27	13.37	7.06	1.67	391.00	284.83	46.32	10.97	0.94	
	Listening	7769	21	16.03	4.68	1.16	416.00	354.10	51.45	12.79	0.94	
1.2	Speaking	7769	24	18.63	5.41	1.16	422.00	358.48	54.26	11.66	0.95	
1-2	Reading	7769	42	29.43	10.36	1.96	396.00	329.26	42.09	7.97	0.96	
	Writing	7769	27	19.13	6.31	1.48	404.00	322.17	50.03	11.76	0.94	
	Listening	9140	18	11.94	3.81	1.75	928.00	702.33	137.76	63.43	0.79	
25	Speaking	9140	24	19.92	4.73	1.57	934.00	791.87	138.25	45.84	0.89	
5-5	Reading	9140	20	12.47	4.69	1.87	922.00	650.65	154.93	61.89	0.84	
	Writing	9140	17	9.27	3.39	1.80	954.00	620.01	128.76	68.65	0.72	
	Listening	6835	18	13.47	3.73	1.58	938.00	800.53	132.37	56.14	0.82	
6 9	Speaking	6835	24	20.53	5.17	1.38	940.00	836.21	133.90	35.74	0.93	
0-8	Reading	6835	20	13.25	4.85	1.78	928.00	702.14	164.87	60.49	0.87	
	Writing	6835	17	11.04	3.65	1.60	956.00	717.08	140.10	61.40	0.81	
	Listening	6456	20	14.52	4.66	1.71	940.00	805.07	132.85	48.85	0.86	
0.12	Speaking	6456	24	19.56	6.12	1.53	936.00	824.63	138.30	34.50	0.94	
9-12	Reading	6456	20	12.29	4.67	1.93	933.00	702.89	143.21	59.20	0.83	
	Writing	6456	20	11.74	4.57	1.90	993.00	706.61	142.32	59.28	0.83	

		Performance Standard Cut Scores Full English								
					Full English					
		Beginners	Intermediate	Advanced	Proficiency					
Grade K Listening	Scaled Score	248	300	355	399					
Grade K Speaking	Scaled Score	255	300	349	394					
Grade K Reading	Scaled Score	270	300	338	359					
Grade K Writing	Scaled Score	251	300	328	375					
Grade 1–2 Listening	Scaled Score	254	300	348	382					
Grade 1–2 Speaking	Scaled Score	266	300	344	388					
Grade 1–2 Reading	Scaled Score	262	300	328	364					
Grade 1–2 Writing	Scaled Score	245	300	329	369					
Grade 3–5 Listening	Scaled Score	450	544	645	725					
Grade 3–5 Speaking	Scaled Score	450	547	668	809					
Grade 3–5 Reading	Scaled Score	450	580	648	770					
Grade 3–5 Writing	Scaled Score	450	577	669	785					
Grade 6–8 Listening	Scaled Score	554	626	718	806					
Grade 6–8 Speaking	Scaled Score	458	611	719	825					
Grade 6–8 Reading	Scaled Score	460	612	690	829					
Grade 6–8 Writing	Scaled Score	553	653	722	894					
Grade 9–12 Listening	Scaled Score	556	632	729	850					
Grade 9–12 Speaking	Scaled Score	570	650	765	850					
Grade 9–12 Reading	Scaled Score	545	630	718	850					
Grade 9–12 Writing	Scaled Score	509	631	719	850					

### **OTELA Cut Score Points for All Performance Standards**

	Percentage of Students at Each Performance Level								
Test Grade Cluster/	Pre-				Full English				
Domain	functional	Beginners	Intermediate	Advanced	Proficiency				
Grade K Listening	4.98	16.18	33.07	26.97	18.80				
Grade K Speaking	9.67	15.96	30.49	24.85	19.02				
Grade K Reading	10.23	25.17	47.64	9.79	7.17				
Grade K Writing	19.83	39.16	23.76	12.71	4.54				
Grade K Comprehension	10.16	24.88	48.06	9.89	7.02				
Grade K Production	18.20	37.95	26.85	12.63	4.37				
Grade K Composite	18.32	38.29	30.56	9.52	3.30				
Grade 1-2 Listening	3.21	10.49	24.60	24.16	37.55				
Grade 1-2 Speaking	4.75	7.93	22.31	27.16	37.86				
Grade 1-2 Reading	5.51	16.36	24.86	30.07	23.21				
Grade 1-2 Writing	6.35	21.77	20.93	29.82	21.14				
Grade 1-2 Comprehension	5.35	15.92	25.67	30.20	22.86				
Grade 1-2 Production	5.92	20.27	23.44	29.84	20.53				
Grade 1-2 Composite	6.92	21.19	27.21	28.74	15.94				
Grade 3-5 Listening	3.68	10.77	19.47	26.24	39.85				
Grade 3-5 Speaking	2.30	3.13	12.00	33.80	48.77				
Grade 3-5 Reading	9.60	24.35	11.63	30.33	24.09				
Grade 3-5 Writing	9.27	29.97	33.63	17.12	10.01				
Grade 3-5 Comprehension	7.93	23.92	14.49	30.30	23.37				
Grade 3-5 Production	5.66	23.81	43.45	17.23	9.86				
Grade 3-5 Composite	9.75	25.88	31.29	26.09	6.99				
Grade 6-8 Listening	6.58	5.66	13.09	15.03	59.63				
Grade 6-8 Speaking	2.37	5.35	9.44	20.54	62.30				
Grade 6-8 Reading	7.99	22.00	16.05	29.70	24.26				
Grade 6-8 Writing	13.86	14.69	17.83	37.86	15.76				

Percentage of Students at Each Performance Level

	Pe	erformance Lev	vel		
Test Grade Cluster/ Domain	Pre- functional	Beginners	Intermediate	Advanced	Full English Proficiency
Grade 6-8 Comprehension	6.70	19.34	20.25	29.69	24.02
Grade 6-8 Production	9.10	13.86	23.54	37.86	15.64
Grade 6-8 Composite	10.87	17.29	25.57	35.67	10.59
Grade 9-12 Listening	5.48	9.00	11.00	21.65	52.87
Grade 9-12 Speaking	6.71	5.70	15.29	15.20	57.11
Grade 9-12 Reading	13.43	17.43	24.97	21.14	23.03
Grade 9-12 Writing	8.36	22.68	20.24	32.31	16.40
Grade 9-12 Comprehension	10.66	17.43	28.00	21.07	22.85
Grade 9-12 Production	6.91	17.97	26.81	32.45	15.86
Grade 9-12 Composite	11.86	19.69	29.06	28.36	11.03

## Equating and Scaling: How Raw Scores Are Converted into Scaled Scores

#### **Test Form Construction**

The Ohio Test of English Language Acquisition (OTELA) is based on the English Language Development Assessment (ELDA) developed under the direction of a consortium of 18 member states of the LEP State Collaborative on Assessment and Student Standards (LEP-SCASS) and the Council of Chief State School Officers. The ELDA was designed to allow states to meet federal requirements under NCLB concerning the annual assessment of LEP students regarding their acquisition of and progress toward developing English language proficiency in listening, speaking, reading, and writing.

The OTELA is a battery of tests designed to allow schools to measure progress in the acquisition of English language proficiency skills among non-native English-speaking students. The battery consists of separate tests for listening, speaking, reading, and writing, at each of five grade clusters: K, 1–2, 3–5, 6–8 and 9–12. The tests are aligned with Ohio's English language proficiency standards and were constructed to provide content coverage across three academic topic areas (English Language Arts; Mathematics, Science and Technology; and Social Studies), and one non-academic topic area, School-Environmental, which is related to aspects of the school environment such as extracurricular activities, student health, homework, classroom management, and lunchtime. Although the OTELA tests measure language skills with content drawn from age-appropriate curricular and non-curricular sources, they are not tests of academic content. Students do not need any external or prior content-related knowledge to respond to the test questions.

To measure a wide range of English language proficiency, the full-length ELDA includes many items and requires substantial test administration time. Although administration of the ELDA test battery is not officially timed, general guidelines indicate approximately four hours of test administration time. In addition, most students to whom the ELDA was administered scored in the upper ranges of the raw score distribution. These performance results indicated that the ELDA operational forms could be shortened substantially by eliminating the easiest items in the operational item bank while maintaining a proportional representation of items across content standards within each domain.

OTELA items were selected on the basis of their psychometric properties, contribution to measurement at key points on the scale (such as the intermediate cut score) and content coverage. When, for example, the easiest items within a domain proved to be concentrated within specific content standards, the Ohio Department of Education (ODE) opted to maintain breadth of content coverage, rather than to simply increase form difficulty. In addition, although a primary goal was to reduce test length as much as possible, estimated form reliabilities were used to determine the appropriate number of items to include in each test form.

#### **Common Item Equating**

*Grade clusters 3–5, 6–8, and 9–12.* Following the first operational administration of grades 3–12 ELDA forms in 2005, items included in the first operational test forms were recalibrated and the resulting item parameter estimates serve as the reference scales for ELDA. All subsequent grades 3–12 ELDA test forms are linked to these scales.

Because the first set of operational forms were constructed to include a set of common items between adjacent grade clusters, the grades 3–5, 6–8 and 9–12 forms were jointly calibrated in a single Winsteps run for each domain, resulting in a common, vertically linked scale across grade clusters for each domain. For each Winsteps run, the mean of the item difficulty parameters was fixed to zero so that the average difficulty for all items across grade clusters was equal to zero within each domain for the first operational form.

For the 2005 field test, a common item design was used to allow common item equating across field-test forms and the first operational form. Following the common item design of the field test, items were jointly calibrated in a single Winsteps run for each domain and grade-cluster combination. Because all of the 2005 ELDA field-test forms shared items in common with operational Form 1, a common item equating method was used to link the field-test items to the ELDA operational Form 1 scale. For each field-test form within each grade cluster, shared items were fixed to their operational Form 1 parameter estimates, while the remaining items were freely estimated. This placed all the field-test items on the operational Form 1 scale.

In addition, a small subset of items were field tested in 2004 but were not included in the 2005 operational forms. These items were also placed on the 2005 operational ELDA scale. Because all items in the 2005 operational test came from the 2004 field-test item pool, the 2005 operational test items were used as linking items. The mean-mean procedure was used to find the linking constant. To ensure that the final set of anchor items (i.e., common items) was free of item parameter drift, a stepwise deletion procedure was used to select anchor items and calculate the linking constant needed to bring the field test items onto the reference scale defined by the first operational administration. Following this procedure, a linking constant was calculated, using all anchor items, and then applied the linking constant to bring the items back to the reference scale. Anchor item parameter estimates were then examined to determine whether the difference between any adjusted or linked parameter estimates and the reference between its linked and reference item parameter estimates, provided the difference was greater than .3. We then computed and applied a new linking constant to the test items and again examined the parameter estimates for the remaining anchor items to determine whether any exceeded the .3 tolerance level. We repeated this process until all remaining anchor items met the tolerance-level specifications. The linking constant was computed on the basis of this final anchor item set, and then applied to the 2004 ELDA field-test item parameters.

The result of these analyses was to place all items in each of the grade 3-12 ELDA domain item banks on common scale defined by the first operational administration.

*Grade clusters K and 1–2.* Items in the grades K and 1–2 OTELA forms were calibrated independently of the items in the grade 3–12 scales and are not reported on the vertical scale used to report scores on the grades 3–12 OTELA tests. A large proportion of items in the listening and speaking tests are common across the grades K and 1–2 test forms, while item overlap between the grades K and 1–2 reading and writing test forms is minimal. Consistent with this perspective, item difficulties for the kindergarten and grade 1–2 OTELA test forms were calibrated following two distinct strategies. Parameters for all OTELA kindergarten and grade 1–2 items were estimated using Masters' partial credit model, an extension of the Rasch model for polytomous items. Student item scores were obtained from the spring 2006 operational administration of the OTELA. For the reading and writing assessments, items in each of the grades K and 1–2 operational test forms were calibrated in separate Winsteps runs. For the listening and speaking items, parameters for items in both the grades K and 1–2 forms were estimated simultaneously in a joint calibration. Once the listening and speaking items were calibrated, the resulting cross-grade item parameter estimates were used to generate form-specific raw score to theta scale conversion tables.

Reporting scales for the grades K and 1–2 OTELA forms were established by setting the "intermediate" or level 3, performance standard for each of the assessments to be 300. Therefore, for both the grades K and 1–2 assessments, and across the four English language domains assessed, a score of 300 indicates attainment of an intermediate level of English language proficiency. The standard deviation of the scale was set to 15.

#### **Performance Standards**

The OTELA is designed to provide student performance-level assessment results that are fully comparable with those from the ELDA. To achieve this goal, the OTELA uses the same performance standards adopted by the LEP-SCASS for the ELDA. In the process of adopting ELDA performance standards for the OTELA, ODE, in consultation with the Ohio LEP Advisory Committee, elected to revise one ELDA performance level cut score. In the ELDA performance standards for writing, students in the grade 3–5 cluster must substantially outperform students in both the 6–8 and 9–12 grade clusters to achieve Full English Proficiency. To address this issue, a linear regression approach was used to identify a cut score for Full English Proficiency at the grade 3–5 cluster from the cut score identified for Beginning, Intermediate, and Advanced performance levels on the grade 3–5 writing assessment. This analysis identified a cut score of 2.08 (in the theta metric; 867 on the ELDA reporting scale) for the Full English Proficiency cut score at the 3–5 grade cluster. AIR submitted the cut score and estimated impact data for the revised performance standard to the Ohio LEP Advisory Committee for their consideration. The Ohio LEP Advisory Committee recommended that ODE adopt the revised performance standard, which ODE has done.

	Scaled Scores Corresponding to Raw Score Points										
	Grade K	Grade K	Grade K	Grade K	Grade 1–2	Grade 1–2	Grade 1–2	Grade 1–2			
<b>Raw Score</b>	Listening	Speaking	Reading	Writing	Listening	Speaking	Reading	Writing			
0	195	187	213	175	178	183	198	167			
1	210	202	226	190	193	198	210	183			
2	225	216	240	205	208	213	222	198			
3	237	227	248	216	220	224	230	208			
4	248	236	254	226	231	233	236	217			
5	260	246	259	235	242	242	241	224			
6	272	255	263	243	254	250	246	231			
7	282	264	267	251	264	258	250	238			
8	291	272	270	257	273	266	254	245			
9	300	279	273	263	282	273	258	251			
10	309	286	276	269	291	280	262	258			
11	318	293	279	275	300	286	265	264			
12	327	300	282	280	309	293	269	270			
13	336	307	284	285	319	300	272	276			
14	346	315	287	290	328	308	275	282			
15	355	323	289	295	338	316	279	288			
16	365	331	291	300	348	324	282	294			
17	375	340	294	305	358	334	285	300			
18	386	349	296	310	369	344	288	307			
19	399	359	298	316	382	354	291	314			
20	416	369	300	322	399	364	294	321			
21	432	380	302	328	416	376	297	329			
22		394	304	334		388	300	337			
23		411	306	341		405	303	345			
24		429	308	349		422	306	356			
25			311	359			309	369			
26			313	375			312	386			

Raw Score to Scaled Score Conversion Table—Grades K-2

			Scaled Scor	res Corresp	onding to Ra	aw Score Poi	nts	
	Grade K	Grade K	Grade K	Grade K	Grade 1–2	Grade 1–2	Grade 1–2	Grade 1–2
<b>Raw Score</b>	Listening	Speaking	Reading	Writing	Listening	Speaking	Reading	Writing
27			315	391			315	404
28			317				318	
29			319				321	
30			321				325	
31			324				328	
32			326				332	
33			329				335	
34			332				339	
35			335				343	
36			338				348	
37			342				352	
38			346				358	
39			352				364	
40			359				372	
41			371				384	
42			383				396	

	Scaled Scores Corresponding to Raw Score Points											
	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade	Grade
Raw	3–5	3–5	3–5	3–5	6–8	6–8	6–8	6-8	9–12	9–12	9–12	9–12
Score	Listening	Speaking	Reading	Writing	Listening	Speaking	Reading	Writing	Listening	Speaking	Reading	Writing
0	162	207	150	198	178	290	151	243	222	332	169	268
1	238	277	189	277	272	352	197	319	313	394	261	341
2	331	347	282	355	366	415	292	395	403	455	354	415
3	390	390	341	405	426	453	353	442	460	492	411	460
4	435	423	386	444	473	481	399	479	503	520	455	494
5	473	449	423	478	512	505	438	511	538	543	491	522
6	507	472	456	509	547	525	473	540	570	563	523	547
7	539	492	486	540	580	544	504	569	599	581	552	571
8	569	511	514	570	612	562	534	598	626	598	580	594
9	598	529	541	601	642	578	563	628	652	614	606	617
10	627	546	568	633	673	594	591	660	677	630	632	641
11	657	563	595	668	705	610	619	695	703	645	658	667
12	689	581	622	707	739	625	648	735	729	660	685	694
13	723	598	650	754	775	641	677	782	756	675	712	724
14	761	617	680	815	816	657	709	840	785	691	742	758
15	806	636	713	894	864	673	743	899	817	706	774	797
16	865	656	751	924	909	691	782	927	853	722	811	845
17	914	678	796	954	923	710	828	956	896	739	855	897
18	928	702	855		938	731	888		913	758	898	920
19		728	904			755	911		927	778	915	956
20		759	922			783	928		940	802	933	993
21		796				818				831		
22		845				865				871		
23		907				914				913		
24		934				940				936		

Raw Score to Scaled Score Conversion Table—Grades 3–12